

(19)日本国特許庁 (JP)

(12)公開特許公報 (A)

(11)特許出願公開番号

特開2000-81897

(P2000-81897A)

(43)公開日 平成12年3月21日(2000.3.21)

(51)Int.Cl.<sup>7</sup>

識別記号

FI

テーマコード (参考)

G10L 19/00

G10L 9/18

J 5D045

G09B 19/06

G09B 19/06

9A001

G10L 19/02

G10L 7/04

G

審査請求 未請求 請求項の数12 O L (全18頁)

(21)出願番号

特願平10-249672

(22)出願日

平成10年9月3日(1998.9.3)

(71)出願人 000104179

カナース・データー株式会社

東京都千代田区東神田1丁目10番7号

(72)発明者 関口 博司

東京都千代田区東神田一丁目10番7号 カ

ナース・データー株式会社内

(74)代理人 100088155

弁理士 長谷川 芳樹 (外3名)

Fターム(参考) 5D045 DB01

9A001 BB02 BB03 DD15 EE02 EE05

FF03 GG03 HH18 JJ33 JJ74

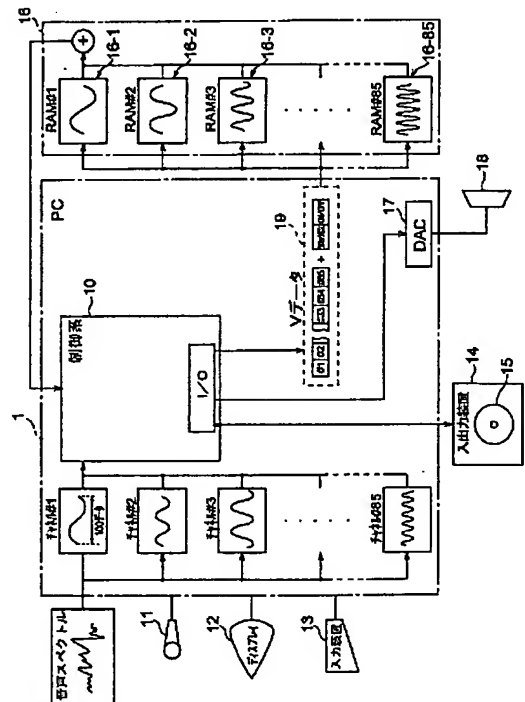
KK09

(54)【発明の名称】 音声情報の記録方法、音声情報記録媒体、並びに音声情報の再生方法及び再生装置

(57)【要約】

【課題】 日本人英語学習者のヒヤリング練習用の音声として、任意の部分において元の音声情報自体の周波数成分を変えることなく再生時間を伸長・短縮及び強調・減衰された音声情報を提供する。

【解決手段】 この発明に係る記録方法は、第1周期でサンプリングされた第1音声情報列を複数の周波数成分に分割し、これら各周波数成分について、第2周期で順次抽出された振幅情報列に対して所定部分の振幅及び所定部分に波形数が変更された正弦波データを生成し、これら各周波数成分に相当する正弦波データを加算して合成された第2音声情報列を所定の記録媒体に記録する。これにより、得られた記録媒体には、周波数を変えることなく任意の部分で伸長・短縮及び強調・減衰された音声情報列が記録される。



## 【特許請求の範囲】

【請求項1】 音声情報を所定の記録媒体に記録するための音声情報の記録方法であって、

第1周期でサンプリングされた第1音声情報列を複数の周波数成分に分割し、

前記複数の周波数成分について、少なくとも1波形分以上に相当する第2周期で抽出された振幅情報からなる振幅情報列における1又は2以上の所定部分が選択的に編集された修正振幅情報列を生成し、

前記複数の周波数成分おのおのの修正振幅情報列のうち、各周波数成分間で互に対応している同じタイミングで抽出された振幅情報からなる各情報成分群と、これら各情報成分群ごとに用意された、前記第1周期を基準にして音声再生時間の伸長あるいは短縮を指示するための制御情報とからなるVデータを生成し、  
前記複数の周波数成分おのおのについて、前記生成されたVデータで与えられる振幅を有するとともに前記第1周期のデータ間隔を有する正弦波データであって、前記Vデータに含まれる制御情報で指示された再生時間に相当する波数の正弦波データを生成し、

前記複数の周波数成分おのおのについて生成された正弦波データを順次加算することにより得られる、前記第1周期の第2音声情報列を所定の記録媒体に記録する音声情報の記録方法。

【請求項2】 前記複数の周波数成分おのおのの振幅情報列間で、互に対応している選択された部分の振幅情報により与えられる振幅値をそれぞれ変更することにより、前記複数の周波数成分おのおのの修正振幅情報列を生成することを特徴とする請求項1記載の音声情報の記録方法。

【請求項3】 前記複数の周波数成分おのおのについて、生成される前記正弦波データの各振幅は、前記修正振幅情報列の互いに隣接した各振幅情報間の直線補間により得られた値により決定されることを特徴とする請求項1又は2記載の音声情報の記録方法。

【請求項4】 前記制御情報は、前記複数の周波数成分全体を高音方向あるいは低音方向にシフトした状態で再生させるための周波数シフト指示情報を含むことを特徴とする請求項1～3のいずれか一項記載の音声情報の記録方法。

【請求項5】 前記第1音声情報列は、所定の音声再生手段で再生出力されるべき単語列から構成された1又は2以上の文に対応する音声情報列であって、発音の節目でそれぞれ分割された情報ごとに可変長の区画に区分された状態で前記所定の記録媒体に記録されることを特徴とする請求項1～4のいずれか一項記載の音声情報の記録方法。

【請求項6】 前記第2音声情報列は、前記第1音声情報列の区画に対応して分割された区画ごとに前記所定の記録媒体に記録されており、さらに該記録媒体には、該

第1音声情報列と該第2音声情報列とを前記所定の音声再生手段で切替え再生すべく、切替え可能な各区画を当該所定の記録媒体における該各区画の記録位置で示す記録位置識別情報が記録されることを特徴とする請求項5記載の音声情報の記録方法。

【請求項7】 前記請求項1記載の音声情報の記録方法により第2音声情報列が記録された音声情報記録媒体。

【請求項8】 所定周期でサンプリングされた第1音声情報列を構成する各周波数成分について、各周波数成分間で互に対応している1又は2以上の部分に対し、少なくとも振幅が変更されるかあるいは波形数を変更されることにより、選択的に振幅及び再生時間が編集された第2音声情報列が少なくとも記録された音声情報記録媒体。

【請求項9】 前記第1音声情報列は、所定の音声再生手段で再生出力されるべき単語列から構成された1又は2以上の文に対応する音声情報列であって、発音の節目でそれぞれ分割された情報ごとに可変長の区画に区分された状態で記録されている特徴とする請求項7又は8記載の音声情報記録媒体。

【請求項10】 前記第2音声情報列は、前記第1音声情報列の区画に対応して分割された区画ごとに記録されており、さらに該第1音声情報列と該第2音声情報列とを前記所定の音声再生手段で切替え再生すべく、切替え可能な各区画を当該所定の記録媒体における該各区画の記録位置で示す記録位置識別情報が記録されたことを特徴とする請求項9記載の音声情報記録媒体。

【請求項11】 所定の記録媒体に予め記録されている音声情報列を再生するための音声情報の再生方法において、

前記記録媒体は、所定の音声再生手段で再生出力されるべき単語列から構成された1又は2以上の文に対応する音声情報列であって、発音の節目でそれぞれ分割された情報ごとに可変長の区画に区分された第1音声情報列と、該第1音声情報列の区画に対応して分割された音声情報列であって、該第1音声情報列を構成する各周波数成分について、各周波数成分間で互に対応している1又は2以上の部分に対し、少なくとも振幅が変更されるかあるいは波形数を変更されることにより、選択的に振幅強度及び再生時間が変更された第2音声情報列と、該第1音声情報列と該第2音声情報列とを前記所定の音声再生手段で切替え再生すべく、切替え可能な各区画を当該所定の記録媒体における該各区画の記録位置で示す記録位置識別情報とを少なくとも含み、

前記第1音声情報列の再生中あるいは中断の後に入力された前記第2音声情報列の再生命令に対し、再生中の前記第1音声情報列中の区画に対応する前記第2音声情報列中の区画の音声情報列を前記記録位置識別情報に基づいて前記記録媒体から読み出し、該読み出された音声情報列を前記所定の音声再生手段により再生する第1ステ

ップと、  
前記第2音声情報列の再生中あるいは中断の後に入力された前記第1音声情報列の再生命令に対し、再生中の前記第2音声情報列中の区画に対応する前記第1音声情報列中の区画の音声情報列を前記記録位置識別情報に基づいて前記記録媒体から読み出し、該読み出された音声情報列を前記所定の音声再生手段により再生する第2ステップと、を少なくとも備えた音声情報の再生方法。

【請求項12】 前記請求項11記載の音声情報の再生方法を実施するための音声情報列の再生装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、CD-ROM、MD、MO等の円盤状記録媒体やDAT等のテープ状記録媒体に音声情報を記録する音声情報の記録方法、該音声情報が記録された音声情報記録媒体、並びに該音声情報記録媒体に記録された音声情報列を読み出し再生するための音声情報の再生方法及び再生装置に関するものである。

【0002】

【従来の技術】従来から、英会話等の語学の独習用、詩吟の練習用、法律の独習用、その他の目的のために、カセットテープ等の記録媒体に音声情報が記録された教材が種々提供されている。ここで、英会話の独習用の教材を例に説明すると、従来の主な記録媒体は、例えば一連の英語の発声（音声情報）が記録されたカセットテープ（又はレコード）であり、学習者はこのテープ教材とテキストとを組み合わせて使用していた。なお、このような教材には、初級用から上級用まで種々のレベルが用意されている。

【0003】また、日本国特許第2581700号には、複数の区画に区分された上級者学習用に適した音声情報列（ナチュラルスピードの発声音）が記録された第1領域と、これら各区画に対応した等価な区画からなる初級者学習用に適した音声情報列（はっきりとした発声音）が記録された第2領域と、該上級者学習用及び初級者学習用の各音声情報列の対応する各区画の関係を、これら音声情報列の各区画の記録媒体における記録位置で示す情報が記録された第3領域とを、少なくとも備えたCD-ROM等の情報記録媒体、及びこのような構造を備えた情報記録媒体の対応する区画間での切替え再生等を含む再生方法が提案されている。

【0004】

【発明が解決しようとする課題】上述のように、日本国特許第2581700号の情報記録媒体には、該媒体上の第1領域にネイティブスピーカーの発声音が記録され、また第2領域に言語上は同一の意味で遅延した発音で構成された音声情報列が記録されている。したがって、第1領域に記録された音声情報列が再生されている

最中に再生音を聞き取れなかった場合、第2領域に記録された同一内容の音声情報列（第1音声情報列の再生中の区画と第2音声情報列の再生すべき区画との対応は第3領域に記録されている）を切替えて再生することにより、学習者は聞き取れなかった音声の意味を理解することができる。

【0005】しかしながら、英語学習者は上述のように第2領域に記録されている情報を聞くことにより第1領域に記録された情報を理解することはできても、依然として該第1領域に記録されている情報、取り分け聞き取れない音は単に繰り返し聞いただけでは聞き取れるようにはならない。日本人英語学習者の場合、日本語にない音素特に子音の聞き取りが苦手であり、ネイティブスピーカーとの会話に支障をきたしていることは周知である。

【0006】この発明は聞き取り難い部分が学習者にとって聞き取りやすいように予め編集された音声の間かせることで、元の音声に対するヒヤリング能力を向上させる技術に関し、英語学習者のヒヤリング練習用の音声情報として、元の音声情報自体の周波数成分を変えることなく選択的に周波数成分の振幅、再生時間が編集された音声情報の記録方法、音声情報記録媒体、並びに音声情報の再生方法及び再生装置を提供することを目的としている。

【0007】

【課題を解決するための手段】この発明は、ヒヤリング練習用の音声として、取り込まれた音声情報列の周波数成分を変えることなく、該音声情報列の所望の部分の強調あるいは減衰させたり、また再生時間を部分的に伸長あるいは短縮させた音声情報列を新たに生成、記録、再生する技術に関するものである。この発明では、再生される音声情報の音質を変えないため、サンプリングされた音声情報に対してではなく、該音声情報の各周波数成分に対して所望の編集を行い、これら編集された周波数成分を合成して新たな音声情報列を得ている。この構成により、日本人英語学習者にとって聞き取り難い部分が選択的に強調及び／又は伸長されたヒヤリング練習用の音声情報の提供を可能にする。また、上級者がヒヤリング能力のさらなる向上を望む場合には、逆に音声を選択的に減衰させたり再生時間が短縮された音声情報の提供を可能にする。

【0008】具体的にこの発明に係る音声情報の記録方法は、第1周期（例えば音楽CDの音響クロック44.1KHz）でサンプリングされた第1音声情報列を複数の周波数成分（以下、チャンネルという）に分割し、第2周期（例えば1波形を形成するために必要なデータ数に相当）で各チャンネルごとにその振幅情報を得る。なお、この振幅情報は第1音声情報列の例えば100データ分に相当する波形の振幅変化量で与えられ、もし100データ分で1波形が形成されない場合には1波形できるデ

一タ数に増やして(第2周期を長くして)抽出される。  
なお、この第2周期は規則性のある周期であればよい。

【0009】さらに、このように得られた各チャンネルの振幅情報列(各チャンネルごとに第2周期で抽出された振幅情報の列)に対してそれぞれ振幅情報を選択的に変更するよう編集された複数の修正振幅情報列が生成される。この複数の修正振幅情報列は、それぞれ各周波数成分に対応したチャンネルごとに求められる。そして、各チャンネルに対応した修正振幅情報列間で、互に対応している同じタイミングで抽出された振幅情報からなる各情報成分群と、これら各情報成分群ごとに用意される、第1周期を基準にして音声再生時間の伸長あるいは短縮を指示するための制御情報とからなるVデータが生成される。

【0010】続いて、上記第2周期のデータとして生成されたVデータから、該Vデータにより与えられる振幅(修正後の値)を有するとともに第1周期のデータ間隔を有する、各チャンネルに相当する正弦波データであって、上記制御情報で指示された再生時間に相当する波数の正弦波データがそれぞれ生成される。このように各チャンネルごとに生成された正弦波データは順次加算されることにより、第1周期のオーディオデータ(第2音声情報列)が生成される。そして、この生成されたオーディオデータが所定の記録媒体に記録される。

【0011】なお、この発明に係る音声情報の記録方法では、第2周期で抽出された各チャンネルの振幅情報列における各振幅情報に対し、任意の部分で選択的に強調されるか減衰されるよう編集が行われる。すなわち、この発明に係る音声情報の記録方法は、各チャンネルの振幅情報列について、各チャンネル間で互に対応している所定部分の振幅情報から与えられる振幅値をそれぞれ選択的に大きくあるいは小さく設定し直すことにより、修正振幅情報列を生成している。また、この発明に係る音声情報の記録方法では、再生音の不自然な振幅変化を避けるため、各チャンネルについて、生成される正弦波データの各振幅は、修正振幅情報列の互いに隣接した各振幅情報間の直線補間により得られた値により決定されることを特徴としている。

【0012】以上のようにこの発明に係る音声情報の記録方法では、各チャンネルごとに生成された振幅情報列に対して、その任意の部分の振幅を変更するよう構成されており、また、再生時間の伸長・短縮を指示するための制御情報が第2周期で抽出された各チャンネルの振幅情報をまとめた情報成分群ごとに用意されるため、周波数成分を変更することなく、任意の部分において該再生音声の選択的な強調・減衰を可能にするとともに、再生時間の部分的な伸長・短縮も可能にする。

【0013】これは、主として日本人がナチュラル・スピードの英語を単にゆっくり再生して聴けるようにした場合であっても、各周波数成分について単純にかつ一様

に音声再生時間を伸ばしたり短縮したのでは不十分であり、発生音の種類によっては子音部のスペクトルの時間変化が言語上の音として別の音を意味する場合があるからである。例えば、BA(バ)とPA(パ)の発音は、前者のスペクトル変化が速く、後者は遅いだけでスペクトルそのものはほとんど同じ形をしている。したがって、BA(バ)という発音の子音部も含めて時間を伸長するとPA(パ)と聴こえることになる。これを防ぐには子音部の伸長度をBA(バ)と聴こえる限界に留め、母音部のみ望みの音声再生時間に伸長あるいは短縮するようにすれば、BA(バ)のままに聴こえることになる。一方、母音部はいくら伸長あるいは短縮してもその母音のままで聴こえるから望みの長さ(望みの再生時間)に設定できる。一方、日本人には弱すぎて聴き取りにくい小さな子音部の音のところだけを選択的に2倍とか3倍に強調して聴かせることも必要である。母音部も含めて強調したのでは全体が大きくなり過ぎて効果がない。どうしても選択的に強調しなければならない。以上の理由から、この発明に係る音声情報の再生方法は、各チャンネルの振幅情報列も初級者にとって特に聞き取り難い部分を選択的に強調された修正振幅情報列を編集し、さらにこれら各チャンネルごとの修正振幅情報列のうち同じタイミングで生成された振幅情報から構成されるVデータとともに再生時間の伸長を指示する制御情報を順次記録するよう構成されている。逆に、上級者の場合には上述の各発声音の特性を考慮して、所望の部分で再生音声が減衰したり、再生時間が短縮されるよう選択的に音声情報列を編集してもよい。

【0014】さらに、この発明に係る音声情報の記録方法では、男性の音声は上述の記録方法で所定の記録媒体に記録された場合、音声再生時間の伸長を行いながら再生すると、出力される音声の周波数スペクトルは不変であっても感覚的により低い音にシフトしたような錯覚を起す可能性がある。逆に音声再生時間の短縮を行いながら再生すると、感覚的により低い音にシフトしたような錯覚を起す可能性もある。そこで、上記制御情報には、半音分あるいは1音分程度高音方向あるいは低音方向へ周波数成分全体をシフトして再生可能にするための周波数シフト指示情報を含むのが好ましい。

【0015】また、この発明に係る技術は、上述の日本国特許第2581700号に開示された技術と組み合わせることにより、飛躍的な学習効果が期待できる。すなわち、ネイティブスピーカーの発声音を発声の節目で分割した可変長の区画に対応して、任意部分の音声伸長及び/又は強調された音声情報を別途用意することにより、聞き取れなかった音声を繰り返し再生して聞くことができるとともに、係る音声の聞き取り難い部分が強調・伸長された音声を聞くことで、元の音声に対するヒヤリング能力の向上が期待できる。また、上級者にとっては、より積極的に学習効果を向上させるため、区画に区

分されたネーティブスピーカの発声音とともに任意部分の音声短縮及び／又は減衰された音声情報を別途用意することにより、敢えて再生時間を短縮して再生したり、子音部を聞こえにくくする（振幅を小さくする）ことも可能であり、ネーティブスピーカの発声音とを組合わせた学習が可能となる。

【0016】具体的に上記第1音声情報列は、所定の音声再生手段で再生出力されるべき単語列から構成された1又は2以上の文に対応する音声情報列であって、発音の節目でそれぞれ分割された情報ごとに可変長の区画に区分された状態で記録媒体に記録される。これにより、上記第2音声情報列は、第1音声情報列の区画に対応して分割された区画ごとに所定の記録媒体に記録され、さらに該記録媒体には、該第1音声情報列と該第2音声情報列とを所定の音声再生手段で切替え再生すべく、切替え可能な各区画を当該所定の記録媒体における該各区画の記録位置で示す記録位置識別情報が記録される。このように、上記第1音声情報列の分割された各区画と第2音声情報列の各区画間での対応関係を予め記録しておくことにより、所望の1又は2以上の区画を繰り返し再生できるとともに、ナチュラルスピードの再生音と、各学習者のレベルに応じて用意された同一発声音をリアルタイムで切替えながら再生することが可能になる。

【0017】したがって、この発明に係る記録方法により所定の音声情報（波形データではなく、各周波数成分の修正された振幅情報列）が記録された音声情報記録媒体が得られる。

【0018】このような音声情報の記録媒体としては、例えばCD-ROM、MD、MO等の円盤状記録媒体や、DAT等のテープ状記録媒体が適用可能であり、必然的に係る音声情報情報の記録媒体には、第1周期でサンプリングされた第1音声情報列を複数の周波数成分に分割し、これら各周波数成分について、第2周期で順次抽出された振幅情報列に対して所定部分の振幅及び所定部分に波形数が増えられた正弦波データを生成し、これら各周波数成分に相当する正弦波データを加算して合成された第2音声情報列が少なくとも記録されている。すなわち、当該音声記録媒体に記録される第2音声情報列は、所定周期でサンプリングされた第1音声情報列を構成する各周波数成分について、各周波数成分間で互に対応している部分に対し、少なくとも振幅が増えられるかあるいは波形数が増えられることにより、選択的に振幅及び再生時間が編集された第2音声情報列である。

【0019】さらに、この発明に係る音声情報記録媒体は、所定の音声再生手段で再生出力されるべき単語列から構成された1又は2以上の文に対応する音声情報列である上記第1情報列が、発音の節目でそれぞれ分割された情報ごとに可変長の区画に区分された状態で記録されることにより、上述の日本国特許第2581700号に開示された技術と組合わせることが可能である。

【0020】以上のような構成の音声情報記録媒体には、上記第1音声情報列とともに、上記第2音声情報列が、第1音声情報列の区画に対応して分割された区画ごとに記録され、さらに該第1音声情報列と該第2音声情報列とを所定の音声再生手段で切替え再生すべく、切替え可能な各区画を当該所定の記録媒体における該各区画の記録位置で示す記録位置識別情報が記録されているので、このような音声情報記録媒体を用意することにより、この発明に係る音声情報の再生方法及び再生装置は、一方の音声情報列の再生中であっても他方の音声情報列の対応する区画の音声情報列についてリアルタイムの切替え再生が可能になる。

【0021】なお、上述されたこの発明の実施形態には、記録ソフト（上述の記録方法をパーソナルコンピュータ等で実施可能なプログラム、あるいは該プログラムが記録された記録媒体）、専用記録装置、使用マニュアル、あるいはこれらの組合せによる販売、係る音声情報記録媒体単体での販売の他、該音声情報記録媒体、音声情報の再生ソフト（パーソナルコンピュータ等で実効可能なプログラム、あるいは該プログラムを記録した記録媒体を含む）、専用再生装置、使用マニュアル、あるいはこれらの組合せによる販売が考えられる。

【0022】

【発明の実施の形態】以下、この発明の一実施例を図1～図14を用いて説明する。なお、図中同一部分には同一符号を付して重複する説明を省略する。

【0023】この発明は、例えば英語学習者のヒヤリング練習に際し、予め聞き取り難い部分を選択的に強調あるいは減衰させたり、再生時間を伸長あるいは短縮させた音声情報の提供を可能にする技術である。したがって、このように予め編集された音声情報を聞いた学習者にとっては、元の音声に対するヒヤリング能力の向上が期待できる。

【0024】図1は、この発明に係る音声情報の記録動作を概略的に説明するため概念図である。まず、マイク11等により、例えば音楽CDの音響クロック44.1KHz（第1周期）でサンプリングされたネーティブスピーカのナチュラルスピードの音声（第1音声情報）がPC1本体に取り込まれ、一旦ハードディスク等に記録される。そして、取り込まれた音声情報を図2の表に示されたように区分された各チャンネル（周波数成分）に分割するためフィルタリングされる。なお、取り込まれる音声情報の周波数範囲は75Hz～10,000Hz、また、サンプリング周波数は音楽CDの音響クロックに合わせて44.1kHz（22.68μs）とする。分割するチャンネル数は85（7オクターブ+1音）とし、各チャンネル#1～#85の中心周波数（中心f）は平均律（1オクターブ当り12平均律とする）の半音列になるように設定する（77.78Hz（D#）～9,960Hz（D#））。

【0025】以上のように各チャンネル#1～#85にそれぞれ分割されたデータは、その振幅情報が2.268msごと(44.1kHzサンプリングの100データに相当、ただし100データで1波形が形成できない場合にはデータ数を増やす)に抽出される。したがって、この実施形態では、各チャンネル#1～#85における振幅情報のサンプリングレート(第2周期)は441サンプル/s(2.268ms)である。なお、このサンプリングレートは、規則性のある周期であればよく、例えば100データ分取り込んだ次に、120データ分取り込んで処理するなど、これら異なるレートで交互に処理を繰り返すような実施形態であってもよい。

【0026】さらに、PC1の制御系10は、2.268msごとにサンプリングされた各チャンネル#1～#85の振幅情報に対し、種々の編集(ディスプレイ12、及びキーボード、マウス等の入力装置13を介して行うことも可能)を行い、2.268msごとの修正振幅情報群を生成する。そして、各チャンネル#1～#85の修正振幅情報(修正振幅情報群を構成している要素)をそれぞれ1バイト(8ビット)で表現し、さらに2バイトの制御情報を付加して87バイト(85チャンネル×1バイト+2バイト)のVデータ19を生成する。

【0027】なお、修正振幅情報は、各チャンネル#1～#85の振幅情報列(2.268msでサンプリングされた振幅情報)における各振幅情報を、任意の部分で選択的に強調あるいは減衰させるよう編集して得られた情報である。すなわち、各チャンネル#1～#85の振幅情報列について、各チャンネル間で互に対応している所定部分の振幅情報から与えられる振幅値をそれぞれ選択的に大きくあるいは小さく設定し直すことにより、修正振幅情報列は生成される。また、上記制御情報は、上述の編集動作により指示された、各チャンネル#1～#85の周波数成分の再生すべき時間の伸長あるいは短縮を指示する伸長指示情報(1バイト)と各チャンネル#1～#85に相当している周波数成分を低音方向あるいは高音方向に半音又は1音だけ全体的にシフトさせて再生させるか否かを指示する周波数シフト指示情報(1バイト)で構成されている。

【0028】上記伸長指示情報は、1データを何msで再生するかの再生クロック数で表現されている。例えば、この伸長指示情報を再生するクロック数の2分の1で表現すると、50で元の再生時間と同じになり、この情報を100に設定すると44.1kHzのクロックとして200クロックで再生することになり、再生時間を2倍に延ばすことが可能となる(この情報は1バイトで表現されるため、最大で $256 \div 50 = 5.12$ 倍まで再生時間の伸長が可能)。逆にこの情報を25に設定すると44.1kHzのクロックとして50クロックで再生することになり、再生時間を1/2に短縮することが可能となる。また、上記周波数シフト指示情報は、全周

波数成分を低音方向あるいは高音方向にシフトさせる場合にON"1"、シフトさせる必要がない場合にはOFF"0"がセットされる。

【0029】以上のように2.268msでサンプリングされた各チャンネル#1～#85の振幅情報を制御系10が所望の編集を施すことにより生成されたVデータ19に基づいて、新たな音声情報列が生成される。

【0030】なお、上記生成されたVデータ19から新たな音声情報列を生成するためには、各チャンネル#1～#85に相当する波長の正弦波を生成する正弦波生成回路16-1～16-85を有する外部装置16が必要になる。各生成回路16-1～16-85には、各チャンネル#1～#85に対応した周波数の正弦波の基本データが記録されたROMと、生成した正弦波データを一旦記録しておくRAM#1～#85をそれぞれ備えており、これら各回路では、制御系10から送られてきたVデータ19の修正振幅情報に基づいて成形され、かつ制御情報の伸長指示情報で指示された波形数の正弦波データをそれぞれのRAM#1～#85に書込む。なお、この正弦波データを構成するデータ間隔は、サンプリング周波数44.1kHzのデータ間隔22.68μsである。

【0031】そして、これら各生成回路16-1～16-85におけるRAM#1～#85に書込まれている正弦波データが44.1kHzのタイミングで順次読み出され、それぞれ加算されることによりオーディオデータ(音声情報列)が生成される。このオーディオデータは制御系10に送られ、I/Oを介してCD-ROM書込装置等の入出力装置14に制御系10から出力される。この入出力装置14は、制御系10から送られてきた44.1kHzのオーディオデータを例えばCD-ROM等の所定の音声情報記録媒体15に記録する。

【0032】上記各生成回路16-1～16-85で行われる正弦波データの生成では、再生音の不自然な振幅変化を避けるため、各チャンネル#1～#85について、正弦波データの各振幅が、修正振幅情報列の互いに隣接した各振幅情報間の直線補間により得られた値により決定される。また、外部装置16で生成されたオーディオデータはそのままDAC17及びAMPを介してスピーカー18から音声として出力してもよい。さらに、このような音声情報記録媒体15としては、例えばCD-ROM、MD、MO等の円盤状記録媒体や、DAT等のテープ状記録媒体が適用可能である。

【0033】一方、この発明は主として日本人がナチュラル・スピードの英語を単にゆっくり再生して聴けるようする技術に関するものであるが、各周波数成分について単純にかつ一様に音声再生時間を伸ばしたり短縮したのでは不十分である。すなわち、図3は音声スペクトルの基本的な形状を示す図であるが、発生音の種類によっては子音部のスペクトルの時間変化が言語上の音として別の音を意味する場合があるからである。例えば、BA



(バ)とPA(バ)の発音は、前者のスペクトル変化が速く、後者は遅いだけでスペクトルそのものはほとんど同じ形をしている。したがって、BA(バ)という発音の子音部も含めて時間を伸長するとPA(バ)と聴こえることになる。これを防ぐには子音部の伸長度をBA

(バ)と聴こえる限界に留め、母音部のみ望みの音声再生時間に伸長あるいは短縮するようにすれば、BA

(バ)のままに聴こえることになる。一方、母音部はいくら伸長あるいは短縮してもその母音のままで聴こえるから望みの長さ(望みの再生時間)に設定できる。一方、日本人には弱すぎて聴き取りにくい小さな子音部の音のところだけを選択的に2倍とか3倍に強調(振幅を大きくして)して聴かせることも必要である。母音部も含めて強調したのでは全体が大きくなり過ぎて効果がない。どうしても選択的に強調しなければならない。以上の理由から、この発明に係る音声情報の再生方法は、各チャンネルの振幅情報列も初級者にとって特に聞き取り難い部分を選択的に強調された修正振幅情報列を編集し、さらにこれら各チャンネルごとの修正振幅情報列のうち同じタイミングで生成された振幅情報から構成されるVデータとともに再生時間の伸長を指示する制御情報を順次記録するよう構成されている。逆に、上級者の場合には上述の各発声音の特性を考慮して、所望の部分で再生音声が減衰させたり、再生時間が短縮されるよう選択的に音声情報列を編集してもよい。

【0034】さらに、この発明では、男性の音声が上述の記録方法で所定の記録媒体に記録された場合、音声再生時間の伸長及び/又は所望部分の音声強調を行いながら再生すると、出力される音声の周波数スペクトルは不変であっても感覚的により低い音にシフトしたような錯覚を起す可能性がある。逆に音声再生時間の短縮及び/又は所望部分の音声減衰を行いながら再生すると、感覚的により低い音にシフトしたような錯覚を起す可能性もある。そこで、上記制御情報には、半音分あるいは1音分程度低音方向あるいは高音方向へ周波数成分全体をシフトして再生可能にするための周波数シフト指示情報が含まれている。

【0035】次に、この発明は、上述の日本国特許第2581700号に開示されているように、ネイティブスピーカーの音声記録された記録媒体を再生等する技術に好適である。以下、係る技術にこの発明を適用する構成について説明する。

【0036】この発明は、上述の日本国特許第2581700号に開示された技術と組合わせることにより、飛躍的な学習効果が期待できる。すなわち、ネイティブスピーカーの発声音を発音の節目で分割した可変長の区画に対応して、任意部分が選択的に伸長あるいは縮小されたり、強調あるいは減衰された音声情報を別途用意することにより、学習者は聞き取れなかった音声を繰り返し再生して聞くことができるとともに、聞き取り能力を向

上させるべく、再生される音声の聞き取り難い部分が伸長あるいは短縮、強調あるいは減衰された音声としても聞くことが可能になる。

【0037】図4は、この発明に係る音声記録媒体に記録されるべき音声情報列を含む各種情報を概念的に説明するための図である。

【0038】まず、音声情報記録媒体15に記録される第1音声情報列(44.1kHzでサンプリングされた音声情報列)は、映画における出演者の会話、日常生活環境における会話等のように、長さの異なる複数のセンテンス(文)から構成され、また、各センテンス(各会話者の音声情報)の間に、音声再生されていない状況、雑音のみが再生されている状況、音楽(BGM)のみが再生されている状況等のランダムに発生する無音声期間が存在し得る一連の音声情報列である。したがって、第1音声情報列は、所定の音声再生手段で再生出力されるべき複数の単語列から構成された1又は2以上の文に対応する音声情報列であって、当該音声情報記録媒体15の第1領域に、図4に示されたように、発音の節目でそれぞれ分割された音声情報ごとに可変長の区画(以下、セグメントという)に区分された状態で記録される。

【0039】一般にネイティブスピーカーの英会話では、1センテンスは概ね3秒程度で発声されるため、記録されるべき音声情報列を構成するセグメントを決定する発音の節目を各センテンスの間に設定することで、図1(a)、(b)あるいは(d)に示されたように、音声情報列を構成する可変長セグメント621、622、799をそれぞれ構成するのが妥当である。なお、会話中のセンテンスの中には図1(c)に示されたように、極端に短いセンテンスも含まれるが、このセンテンス701も1つのセグメントを構成する。一方、図1(e)に示されたように、極端に長いセンテンスの場合には、接続詞や関係詞等の前が発音の節目となるため、図1(e)に示されたようなセンテンスでは、連続する2つのセグメント801、802で構成するのが妥当である。したがって、記録されるべき音声情報列のセグメントとは、発声上の区切り(息継ぎ位置)又は言語上(文法上)のなんらかの区切りにもとづいて分割された音声情報の記録単位であることを意味する。

【0040】この発明に係る音声情報の記録方法では、まず上述のように第1情報列を分割して得られた各セグメントそれぞれに対し、任意の部分が選択的に編集(各周波数成分の振幅の変更、再生時間の変更)された第2音声情報列を生成する。この第2音声情報列は、具体的には図5に示されたように、各周波数成分について編集するPC1本体と、編集されたオーディオデータ(第2音声情報列)を生成する外部装置16で構成された装置により、所定の音声情報記録媒体15に記録される。

【0041】特に、外部装置16は、図5に示されたよ

うに、オーディオデータを生成するマスターボード165と、各チャネルに対応して設けられた正弦波生成回路16-1~16-85を備えたスレーブボード166で構成されている。マスターボード165は、PC1からのVデータをコントロール信号に従って各生成回路16-1~16-85に供給すべく、タイミングコントローラ171と、FIFO172を備えるとともに、各生成回路16-1~16-85から送られてきた正弦波データ(16ビット)を順次加算し、オーディオデータ(16ビット)を生成する加算器173と、PC1へ送信される該生成されたオーディオデータを一旦格納するバッファとしてのRAM174を備える。なお、図5に示されたマスターボード165は、PC1からの指示で第1音声情報列と新たに編集された第2音声情報列とを音にして何度でもスピーカに出し、耳で聴いて比較できるように、生成されたオーディオデータを直接スピーカ177で再生出力すべく、DAC175及びAMP176が設けられている(音声再生のための構造は図1に示されたようにPC1側に設けられてもよい)。一方、スレーブボード166は、各チャネルに対応して所定の周波数の正弦波をそれぞれ生成する正弦波生成回路16-1~16-85を備えており、これら生成回路16-1~16-85は、正弦波を生成するためのデータが記録されたROMと、一旦生成された正弦波データを格納するバッファとしてのRAM#1~#85をそれぞれ有する。

【0042】なお、マスターボード165とスレーブボード166は、30本の信号バスとGND、Vccの合計32本のバスで接続されており、図中、167で示されたバス群は各生成回路16-1~16-85へVデータを供給するためのVデータ関連バス群であり、168で示されたバス群は各生成回路からマスターボード165へオーディオデータ生成用の正弦波データを送るためのオーディオデータ関連バス群である。

【0043】次に、この発明に係る音声情報の記録方法の、日本国特許第2581700号に開示された技術に適用された実施形態を、図5を参照しながら、図6及び図7のフローチャートを用いて説明する。

【0044】まず、Vデータの生成はPC1側で行われる。すなわち、PC1では、一連の音声情報列(第1音声情報列)が44.1kHz(16bit/データ)をサンプリングし、この第1音声情報列に相当するサンプリングデータを一旦ハードディスクに格納し(ステップST1)、図4に示されたように複数のセグメントに分解する(ステップST2)。

【0045】続いて、分割されたセグメントのうち1セグメントについて、デジタル・バンド・パス・フィルター・プログラムにより、まず第1チャンネル#1のバンド幅(75.57kHz~80.06kHz)の波形情報をメモリーに展開する。この時も44.1kHzのレートに相当するデータ間隔のまま展開する。そして、1

00データごとに平均振幅情報(8ビット)を抽出する(ステップST3)。なお、上述のように第1チャンネル#1の周波数成分について100データで1波形できない場合には1波形できるデータ数に増やして振幅情報を求める。対象セグメントのサンプリングデータが終了するまで、100データ分づつずらして同じ動作を繰り返す。この動作により、対象チャネルである第1チャンネル#1についてデータ間隔2.268msの振幅情報列(1秒当り441個の振幅情報)である。対象チャネルである第1チャンネル#1の振幅情報抽出動作が終了すると(ステップST5)、続いてデジタル・バンド・パス・フィルターにより第2チャンネル#2の周波数を分割して上記ステップST3~ST5の動作を繰り返し、対象チャネルを変更しながら(ステップST7)、第1チャンネル#1~第85チャンネル#85について対象セグメントの振幅情報列が生成される。

【0046】以上の動作は、対象セグメントを変更しながら(ステップST10)、ステップST1でサンプリングされた第1音声情報列を構成するすべてのセグメントが終了するまで行われる(ステップST9)。

【0047】次に、以上のステップST1~ST9が実行されることにより得られた、各セグメントについて85チャンネル分の振幅情報列に対し、PC1側では以下のような編集が行われ、Vデータが生成される(ステップST11)。

【0048】まず、分割されたセグメントごとに生成された85チャンネル分の振幅情報列群を格納先であるハードディスクから呼び出し、モニタ12上に順次その振幅波形を表示する。

【0049】実際の編集作業は、表示された振幅波形の所望の部分を指定して再生時間を指定する(クロック50が基準)。また、必要であれば変更する部分を指定して振幅の変更(表示された震央くを基準にして倍率で設定)を行ったり、低音方向あるいは高音方向への周波数シフト指示を指定する。例えば、セグメントの中の子音部は振幅を2倍、再生時間を1.5倍にする一方、母音部は振幅をそのままにして、再生時間のみ2.5倍にする等、選択的に任意の部分に対して得られた振幅情報列を編集し、新たに各振幅情報が修正された修正振幅情報列を生成する。

【0050】そして、得られた85チャンネル分の修正振幅情報列のうち、各修正振幅情報列間で互に対応している同じタイミングの情報成分をまとめた情報成分群ごとに、上述の再生時間の変更を指示する情報と周波数シフトを指示する情報とからなる制御情報を付加することにより、データ間隔2.268msのVデータが得られる。

【0051】次に、以上のようにPC1側で用意されたVデータ(87バイト/データ)は外部装置16のマスターボード165へ送られ、さらに該マスターボード1



65からデータバスを介してスレーブボード166上の各正弦波生成回路16-1~16-85へ送られる。なお、スレーブボード166は、実際には8回路が搭載された11枚のボード(11枚目のボードには8回路中5回路だけ使用する)で構成されるものとし、それぞれの回路が対応するチャンネルの正弦波データを生成する(ステップST12)。なお、各回路は、正弦波の波形データを収納しているROMが異なることと、対応するチャンネルを指定する7ビットのDIP・SWの設定が異なること以外は全て同じで構成である。

【0052】各チャンネルを受け持つ各回路では、まず、マスターボード165から送られてきた87バイトのVデータのうちヘッダー(2バイト)を共通に受け取る一方、該Vデータのうちの修正振幅情報については対応するチャンネル用の修正振幅情報(1バイト)だけを受け取る。各回路では、44.1kHzの何クロック分で波形を成形し出力するのかを判断するため、受け取ったヘッダー情報の再生時間を調べられる。例えば指示された再生時間が50で与えられた場合には100クロック再生(再生時間は変らない)、110の時は220クロック再生(再生時間は2倍)となる。各回路には受け持つ周波数の正弦波データが44.1kHzで出力された時のデータ間隔でROMに収納されている(ROMのアドレスのゼロ番地からN番地までにその周波数の正弦波が正確にM波収納されている(M、Nは自然数)。各回路中のプロセッサは、1つの正弦波データを作るごと(22.68μsごと)にROMのアドレスを+1していく。そして、N番地の次にはゼロへ戻る。こうすることで、正確な正弦波を不連続点なしに作れる。ただし、上記正弦波データは、受け取った修正振幅情報をそのROMに格納されていた基本データに掛けて1つの正弦波データを生成する。また、各振幅情報は、今回の振幅情報と前回の振幅情報との間を直線補間することにより得られた値とする。

【0053】以上のように、各回路で生成された正弦波データは、周期44.1kHz(22.68μs)でROMを参照し、上記補間で求められた係数を参照されたデータに掛けて出力バッファである各RAM#1~#85へ収納する。

【0054】そして、マスターボード165からのコントロール信号により、各RAM#1~#85に格納された正弦波データが出力バス(16ビット)へ送出するタイミング(22.68μs周期)をもらい、その時だけバスへ送出する。1回路に与えられた時間幅は226ns(22.68μs÷85)となる。一方、マスターボード165側の取込タイミングはクロックと同期信号で与えられる。同期信号から何クロック目かの数は上記DIPスイッチで指定されたチャンネル番号と同じとなる。また、上記Vデータのヘッダー情報に含まれる周波数シフト指示情報がONの時、半音(又は全音)シフト

した周波数の正弦波データが各回路から出力できるよう、各回路に設けられているROMに2種類の波形データを格納しておき、いずれかを選択できるようにする。

【0055】一方、マスターボード165は各回路16-1~16-85で生成された正弦波データを、22.68μs中85データの割合(データ間隔は22.68μs÷85=266ns)で受け取る。実際には、各回路からの正弦波データを受け取りながら加算器173で加算していき、44.1kHzのオーディオデータ(第2音声情報列)を生成する(ステップST13)。生成されたオーディオデータは順次バッファであるRAM174に格納され、PC1へ送られる。

【0056】PC1では、送られてきたオーディオデータを入出力装置14を制御しながら所定の記録媒体15に該オーディオデータを記録していくことにより(ステップST14)、この発明に係る音声情報記録媒体が得られる。

【0057】次に、この発明に係る音声情報記録媒体の、上述の日本国特許第2581700号に開示された技術が適用された各実施形態について説明する。

#### 【0058】音声情報記録媒体に係る第1実施形態

まず、第1実施形態では、少なくとも2種類の音声情報列と記録位置識別情報が記録されている。すなわち、第1音声情報列は例えばネイティブスピーカが自然な速さで話す英語の音声情報からなり、この音声情報列は上述されたように発音の節目(センテンスの終りやセンテンス中の一息つける、発生上あるいは文法上の区切り)で複数の可変長セグメントに分割されている。第2音声情報列は、第1情報列を、上述のように任意の部分が選択的に編集することにより得られた音声情報列であって、第1音声情報列の各セグメントに対応して複数の可変長セグメントに分割されている。また、記録位置識別情報は、少なくとも、第1及び第2音声情報列における各セグメントが、当該音声記録媒体のどの位置に記録されているかを示す情報である。したがって、例えば第1音声情報列のt番目のセグメント”It’s not much of a problem.”に対応する第2音声情報列のセグメント”It’s not much of a problem.”が、媒体のどの位置に記録されているかということは、この記録位置識別情報により認識することができる。

【0059】その結果、第1及び第2音声情報列と記録位置識別情報は互いに無関係に記録されるのではなく、一定の関係をもって記録され、各音声情報列はセグメントを単位として有機的に組み合わせられている。すなわち、第1及び第2音声情報列は互いに対をなしており、これらをセグメントごとに関連させているのが記録位置識別情報である。なお、この実施形態では、記録位置識別情報は当該音声情報記録媒体のディレクトリ領域に記録されており、少なくとも各セグメントの先頭位置に関する情報を含んでいる。

【0060】以上のような構造を備えた音声情報記録媒体（第1実施形態）の再生方法では、記録されたセグメントごとに順番に音声再生が行われるが、特に、この再生方法では、当該音声情報記録媒体に記録された第1音声情報列から第2音声情報列への再生切換え（あるいは第2音声情報列から第1音声情報列への再生切換え）が可能であることを特徴としている。なお、この再生切換え動作は、セグメントを単位として行われる。例えば、第1音声情報列のt番目のセグメントが再生されているときに第2音声情報列の再生指示が入力されると（割込み要求の発生）、記録位置識別情報に基づいて第2音声情報列の対応するt番目のセグメントを読み出し、その対応するセグメントの音声再生が実行される。また逆に、第2音声情報列から第1音声情報列への再生切換えも、上述した再生切換え動作と同様に各セグメント単位で行われる。

【0061】なお、この再生方法では、上述の再生切換え動作の他、リピート再生等の種々の変形が可能である。その代表的なものとして、いわゆる戻し指令がある。すなわち、再生中の停止命令により一時再生を中断した後に戻し指令が入力されたときは、指令された量だけ音声情報の読み出し位置を戻すことによりより操作者の希望に合った音声情報の再生が行われる。

【0062】音声記録記録媒体に係る第2実施形態  
この第2実施形態は、上述された第1実施形態と基本的には同じ構造であるが、上記第1音声情報列及び第2音声情報列の他、第1音声情報列の内容と等価な意味内容であるが別の音声情報であり、例えば単語を区切って話すゆっくりとした速さの英語の音声情報である第3音声情報列を備えていることを特徴としている。また、この第3音声情報列も、複数の可変長セグメントから構成されており、上記記録位置識別情報は、これら第1～第3音声情報列における各セグメント間での記録位置を管理している。したがって、この第2実施形態における音声情報の再生方法は第1実施形態と同様である。

【0063】なお、この実施形態において、重要なことは、上記第1音声情報列と、第3音声情報列はそれぞれ複数の可変長セグメントに区分されているが、互いにセグメントごとにその意味内容が対応していることである。例えば、第1音声情報列のt番目（図4（a）では621番目）のセグメントがネイティブスピーカの話す”It's not much of a problem.”であるときは、第3音声情報列のt番目のセグメントは各単語を区切って話す”It is not much of a problem.”となる。ただし、第2音声情報列と対応した内容でかつ別の音声情報からなるということは、言語上は同一の意味で発声の異なるものであることを示している。

【0064】音声情報記録媒体に係る第3実施形態  
さらに、この発明に係る音声情報記録媒体の第3実施形態について説明する。この第3実施形態に係る音声情報

記録媒体は、第1及び第2音声情報列の他、さらに文法解説等の音声情報列である第4音声情報列が当該音声情報記録媒体に記録されている点が、上述の第1実施形態に係る音声情報記録媒体と異なる。

【0065】ここで重要なことは、上記第3音声情報列は第1及び第2音声情報列の1又は2以上の可変長セグメントをひとまとまりとしたセグメント群に区分されていることである。換言すれば、この第4音声情報列の1つのセグメント群は第1及び第2音声情報列の1又は2以上のセグメントを包含しており、したがって、第4音声情報列の1つのセグメント群は第1及び第2音声情報列の1又は2以上のセグメントと対になっている。特に、この構成は図4（e）に示されたように、1つのセンテンスが複数のセグメントに区分された場合を想定している。

【0066】また、この第3実施形態の音声情報記録媒体において、所定の領域に記録された記録位置識別情報には、上記第4音声情報列の内容の記録位置をもセグメント群ごとに示す情報も含まれている。したがって、第1、第2及び第4音声情報列と記録位置識別情報は互いに一定の関係をもって媒体に記録され、各音声情報列はセグメントあるいはセグメント群を単位として有機的に組み合わせられている。なお、この第3実施形態においても、記録位置識別情報は当該音声情報記録媒体のディレクトリ領域に記録され、各音声情報列におけるセグメントの先頭位置に関する情報を含んでいる。また、この実施形態においても、第1音声情報列の音声情報と等価であって、単語を区切って話すゆっくりとした速さの第3音声情報列をさらに記録してもよい。

【0067】以上のような構造を備えた音声情報記録媒体（第3実施形態）の再生方法は、基本的に上述された第1実施形態の場合と同じであるが、第1及び第2音声情報列間での再生切換えの他、該第1及び第2音声情報列と第4音声情報列との間においても再生切換え動作を行う点が異なる。

【0068】例えば、第1音声情報列の再生中にネイティブスピーカの”It's not much of a problem.”が聴き取れなかったときは、再生中の第1音声情報列から第2音声情報列に再生を切換えることにより、選択的に伸長等の編集が施された音声”It's...not...much...of...a...problem.”を聴くことができる。そして、この日本語の意味や文法を知りたいときは、さらに、再生中の音声情報列から第4音声情報列へ再生を切換えればよい。もちろん、この再生方法においても、上述の第1実施形態に係る音声情報記録媒体の再生方法で説明された戻し指令や停止命令を組み合わせるよう応用できることは言うまでもない。また、この再生方法においても、切換え再生及びリピート再生が可能である。

【0069】音声情報記録媒体に係る第4実施形態  
この発明に係る音声情報記録媒体の第4実施形態は、基

本的に上述の第1実施形態の場合と同様であるが、第1及び第2音声情報列の他、文字情報列が記録されている点が主に異なる。この文字情報列は、第1又は第2音声情報列に対応する内容の文字情報に相当しており、例えばネイティブスピーカが話す英語（音声）に対応する文字情報に相当している。

【0070】この文字情報列も、第1及び第2音声情報列の各セグメントと対応するセグメントに区分されている。また、この第4実施形態に係る音声情報記録媒体においても、記録位置識別情報には、この文字情報列の記録位置を各音声情報列のそれぞれのセグメントごとにそれらの先頭位置に関する情報が含まれ、当該音声情報記録媒体のディレクトリ領域に記録される。したがって、第1及び第2音声情報列と文字情報列はそれぞれセグメント単位で対応することになる。

【0071】なお、この第4実施形態に係る音声情報記録媒体において、上述の第3実施形態における第4音声情報列を記録情報として加えるときは、第1及び第2音声情報列と文字情報列の1又は2以上のセグメントは第3音声情報列の1つのセグメント群にも対応することになる。この構成においても、上記記録位置識別情報には、各セグメントの先頭位置が含まれ、かつ当該音声記録媒体のディレクトリ領域に記録される。そして、上述の第3実施形態と同様に、この第4実施形態でも、第1音声情報列の音声情報と等価であって、単語を区切って話すゆっくりとした速さの第3音声情報列をさらに記録してもよい。

【0072】以上のような構造を備えた音声情報記録媒体（第4実施形態）の再生方法も、基本的に上述の第2実施形態の場合と同様であるが、第1又は第2音声情報列の再生中に文字情報列がディスプレイ表示される点が異なる。

【0073】例えば、第1音声情報列のセグメント“It's not much of a problem.”が再生されているときは、所定の表示部に“It's not much of a problem.”もしくは“It is not much of a problem.”がディスプレイ表示される。なお、この表示については再生中の音声情報列と時間的に完全に同期している必要はなく、文字が少しずつ遅れて表示されたり、あるいは少しずつ先に表示されたりしてもよい。また、この再生方法でも、切換え再生及びリピート再生が可能である。

【0074】次に、この発明に係る音声記録媒体の具体的な構造を、図8～図11を用いて、以下詳細に説明する。

【0075】図8は、この発明に係る音声情報記録媒体の例として、上述の第3実施形態を英会話独習用に適用したときの各音声情報列A、B、Cと、その記録内容を説明するための図である。この図において、音声情報列Aはネイティブスピーカの話す英語の情報列（第1音声情報列）であり、複数のセグメント621、622から

構成されている。音声情報列Bは図6及び図7に示されたフローチャートを用いて説明されたように選択的に該第1情報列の所定部分が伸長等するよう編集された情報列（第2音声情報列）である。また、音声情報列Cは日本語の解説をする情報列（第3音声情報列）であり、この音声情報列Cに含まれるセグメント群は、各音声情報列A、Bの各セグメント621、622にそれぞれ対応している。

【0076】また、図9は、図8に示された態様における1セグメント当りの時間と容量の関係を説明するための表である。この表において、1秒間は6キロバイトの容量に対応している。例えば音声情報列Aのセグメント621では、“It's”の発声時間が0.2秒、その容量が1.2KB（キロバイト）、“not”の発声時間が0.1秒、その容量が0.6KB（キロバイト）、“much of a”の発声時間が0.4秒、その容量が2.4KB（キロバイト）、そして“problem”の発声時間が0.3秒、その容量が1.8KB（キロバイト）であり、セグメント621全体の発声時間は2.0秒、その容量は12KB（キロバイト）となる。

【0077】さらに、図10は、図8及び図9に示された形態におけるディレクトリ領域の記録内容を説明するための表である。この表において、ディレクトリ領域は、1セグメント当り $9 \times 3 = 27$ バイト（B）で構成される。音声情報列A、B、Cはそれぞれ図8の音声情報列A、B、Cに対応している。また、1バイトのCは属性を示し、C=0は音声情報列A、C=64は音声情報列Bであることを意味する。また、C=128、129は音声情報列Cであることを意味し、特にC=129のとき、すなわちビット表現（8ビット（bit））で“10000001”のときは前のセグメントと同じ解説対象であることを示す（音声情報列Cの解説対象となる同じセグメント群に属していることを示し、例えば図4（e）のセグメント801、802の場合が相当する）。

【0078】位置情報のM、S、B（各1バイト）は産業界で標準になっているCD-ROM上の位置を表わすパラメータである。すなわちMは分、Sは秒、Bはブロックをそれぞれ示す。また、1ブロックは2,048バイトであり、75ブロックで1秒分を構成している。したがって、最大数はM=59、S=59、B=74となる。次の2バイトのSBはスタートバイトを示し、その次の3バイトのLLLは各セグメント全体の長さを示している。なお、位置を示すパラメータに分、秒を使う理由はCD-ROMはもともと音楽用として開発されたためであり、始めからの時間として記録位置を表現している。そのためCD-ROMを当該音声情報記録媒体として採用した場合には、この分と秒は再生時の時間とは全く無関係であり、単に記録媒体上の記録位置を表わしている情報にすぎないことになる。

【0079】その結果、例えば音声情報列Aにおけるセグメント621の”It's not much of a problem.”は、0分11秒3ブロックの826バイト目から6,000バイトの長さでネイティブスピーカの話す英語の音声情報が記録され、音声情報列Bにおける対応するセグメントは0分11秒3ブロックの2,026バイト目から17,400バイトの長さで選択的に伸長された上記ネイティブスピーカの英語が記録され、音声情報列Cのセグメント群は0分11秒6ブロックの1,282バイト目から72,000バイトの長さで日本語解説が記録される。なお、621、622等のセグメントナンバーはメモリ上ではなく、そのアドレスに対応している。また、各セグメントの関係を示す記録位置識別情報は、このディレクトリ領域に含まれる。

【0080】さらに具体的には、第10図に示されたディレクトリ領域の記録内容から、当該音声情報記録媒体の0分11秒3ブロックにおける826バイト目から826+6,000-1=6,825バイト目までの領域には、セグメントが621で属性Cが0の音声情報列すなわちネイティブスピーカが話す”It's not much of a problem.”に相当する情報が記録される。また、当該音声情報記録媒体の0分11秒3ブロックにおける2,026バイト目から2,026+17,400-1=19,425バイト目までの領域には、セグメントが621で属性Cが64の音声情報列すなわち選択的に伸長された音声情報が記録される。さらに、当該音声情報記録媒体の0分11秒6ブロックにおける1,282バイト目から1,282+72,000-1=73,281バイト目までの領域には、セグメントが621で属性Cが128の音声情報列すなわち日本語の解説に相当する情報が記録される。

【0081】このように、図10に示されたディレクトリ領域を設ければ、図9に示されたような再生時間及び容量で図8に示された各音声情報列が記録可能である。

【0082】次に、各セグメント621、622に関する情報は、例えば図11(a)に示された可変長セグメントのヘッダ部に記録される。このヘッダ部は、図11(b)に示されたように、先頭から文字情報や画像情報の有無等を示すための1バイト領域(1B)、音声情報列A用に用意された領域であって情報列タイプ(音声情報列A、B等を区別するための情報)を示す1バイトデータ、そのデータ長を示す3バイトデータ、及び予備の1バイトデータから構成された5バイト領域(5B)、音声情報列B用に用意された領域であって情報列タイプを示す1バイトデータ、そのデータ長を示す3バイトデータ、及び予備の1バイトデータから構成された5バイト領域(5B)、音声情報列C用に用意された領域であって情報列タイプを示す1バイトデータ及びそのデータ長を示す3バイトデータから構成された4バイト領域(4B)、文字情報列D用に用意された領域であつ

て情報列タイプを示す1バイトデータ及びそのデータ長を示す3バイトデータから構成された4バイト領域(4B)、同様に文字情報列D用に用意された領域であってアドレスを示す3バイトデータ及びそのデータ長を示す3バイトデータから構成された6バイト領域(6B)、上記第3音声情報列のような他の情報列(タイプE)用に用意された4バイト領域(4B)、及び予備の3バイト領域(3B)からなる、32バイトの領域である。

【0083】次に、図12～図14を用いて、この発明に係る音声情報の再生方法及び装置構成を説明する。

【0084】まず、図12は、この発明に係る音声情報の再生方法実現するための再生装置の全体構成を示す斜視図である。この図からも分かるように、当該音声記録媒体は、例えばポータブルなCDプレーヤ(再生装置本体200)により再生可能なCD-ROMであり、この再生装置本体200はコード接続されたハンドセット80によりリモート制御される。このハンドセット80には少なくとも再生中のセグメント番号を表示する液晶ディスプレイ(LCD)等の表示部210や、各種制御用ボタン群240が設けられている。また、操作者は再生装置本体200で再生された音声情報をイヤホン130を介して聴くことができる。

【0085】また、図13は、図12に示された再生装置の構成を示すブロック図である。この図に示されたように、当該音声情報記録媒体15であるCD-ROMは再生機構205にセットされる。再生機構205はディスクインターフェイス(I/F)30及びバス40を介してCPU50に接続されている。また、バス40にはプログラムを格納するための例えば32キロバイト(KB)のROM60と、ディレクトリや音声情報列を一時的に格納するための例えば256キロバイトのRAM70とが接続されている。さらに、バス40には手動操作のためのハンドセット80との間で情報の授受を行なうハンドセットインターフェイス(I/F)90と、音声出力用のアンプ(AMP)100を介して外部端子110及びハンドセット80に接続されたD/Aコンバータ12に接続されている。なお、ハンドセット80には上述されたようにイヤホン130が接続されている。

【0086】図14(a)、(b)は、それぞれROM60及びRAM70のメモリ割り当て状況を説明するための図である。図14(a)に示されたように、32キロバイトのROM60にはプログラムが格納される。一方、図14(b)に示されたように、RAM70には、 $(50+50)=100$ キロバイトのバッファ(50ブロック分に相当)と、 $(75+75)=150$ キロバイトのディレクトリと、6キロバイト分のシステムエリアが割り当てられる。したがって、RAM70には常時50ブロック分の音声情報列が保持され、かつ150キロバイト $\div 27 \div 5$ 、555セグメント分のディレクトリ(音声情報列Aの部分のみで約30分間に相当)が保持

される。

【0087】なお、上述の具体例では当該音声情報記録媒体としてCD-ROMを用いているが、その代表的なものの容量は552メガバイト(MB)である。CD-ROMではアドレスを表わすのに分、秒、ブロックの単位を用いている。また、1ブロックは2,048バイト、75ブロックは1秒、60秒は1分であるため、該CD-ROMのアドレスの最大の値は59分59秒74ブロックである。逆に、このCD-ROMの容量は、 $2,048 \times 75 \times 60 \times 60 = 552.96$ メガバイトである。このうち、最初から2秒分はCD-ROMのフォーマットとしてユーザは使えないので、正確には最大容量とし552.6528MBとなる。さらに、最初から20秒に相当するところまでディレクトリ領域が割り当てられると、3メガバイトのディレクトリ容量をCD-ROMに確保することができる。

【0088】なお、上述されたこの発明の実施形態には、音声情報の記録ソフト(上述の記録方法をパーソナルコンピュータ等で実施可能なプログラム、あるいは該プログラムが記録された記録媒体)、専用記録装置、使用マニュアル、あるいはこれらの組合わせによる販売、係る音声情報記録媒体単体での販売の他、該音声情報記録媒体、再生ソフト(パーソナルコンピュータ等で実効可能なプログラム、あるいは該プログラムを記録した記録媒体を含む)、専用再生装置、使用マニュアル、あるいはこれらの組合わせによる販売が考えられる。

【0089】

【発明の効果】以上のようにこの発明は、第1周期でサンプリングされた第1音声情報列から分割された複数の周波数成分について、所望の部分に振幅を変更(強調あるいは減衰)したり波数を変更(再生時間を伸長するように増やすあるいは再生時間を短縮のために減らす)することにより、修正された正弦波データを生成し、これら各周波数成分の正弦波データを加算することにより、新たに合成された第2音声情報列を所定の記録媒体に記録する。このように記録された所望の音声情報列は、周波数を変えることなく任意の部分で再生時間を伸長あるいは短縮したり、任意部分の音声が強調あるいは減衰された音声として再生できるという効果がある。

【0090】また、この発明は、日本国特許第2581700号に開示された技術との組合わが可能であり、ネイティブスピーカの発声音を発声の節目で分割した可変長の区画に対応して、任意部分の音声伸長及び/又は強調された音声情報を別途用意することにより、初級学習者は聞き取れなかった音声を繰り返し再生して聞く

ことができるとともに、係る音声の聞き取り難い部分が強調・伸長された音声としても聞くことが可能になるという効果がある。また上級学習者にとっては、任意部分の音声伸縮及び/又は減衰された音声情報を別途用意することにより、ネイティブスピーカの発声音の再生と組合わせて、より積極的な学習が可能になるという効果がある。

【図面の簡単な説明】

【図1】この発明に係る音声情報の記録動作を説明するための概念図である。

【図2】サンプリングされた入力音声情報から分割する各周波数成分(チャンネル)の一例を示す表である。

【図3】音声スペクトルの基本的な形状を説明するための図である。

【図4】この発明に係る音声記録媒体に記録されるべき音声情報列を含む各種情報を概念的に説明するための図である。

【図5】この発明に係る音声情報の記録方法を実現するための周辺装置の全体構成を示す図である。

【図6】この発明に係る音声情報の記録方法を説明するためのフローチャート(その1)である。

【図7】この発明に係る音声情報の記録方法を説明するためのフローチャート(その2)である。

【図8】この発明に係る、英会話独習用に適用された音声記録媒体の各音声情報列と、その記録内容を説明するための図である。

【図9】図8に示された各音声情報列について、各セグメント当りの時間と容量との関係を説明するための表である。

【図10】図8及び図9に示された音声記録媒体におけるディレクトリ領域の記録内容(記録位置識別情報を含む)を説明するための表である。

【図11】この発明に係る音声記録媒体に記録されるべき可変長セグメントの構成を示す図である。

【図12】この発明に係る音声記録媒体の再生方法を実現する再生装置の全体構成を示す斜視図である。

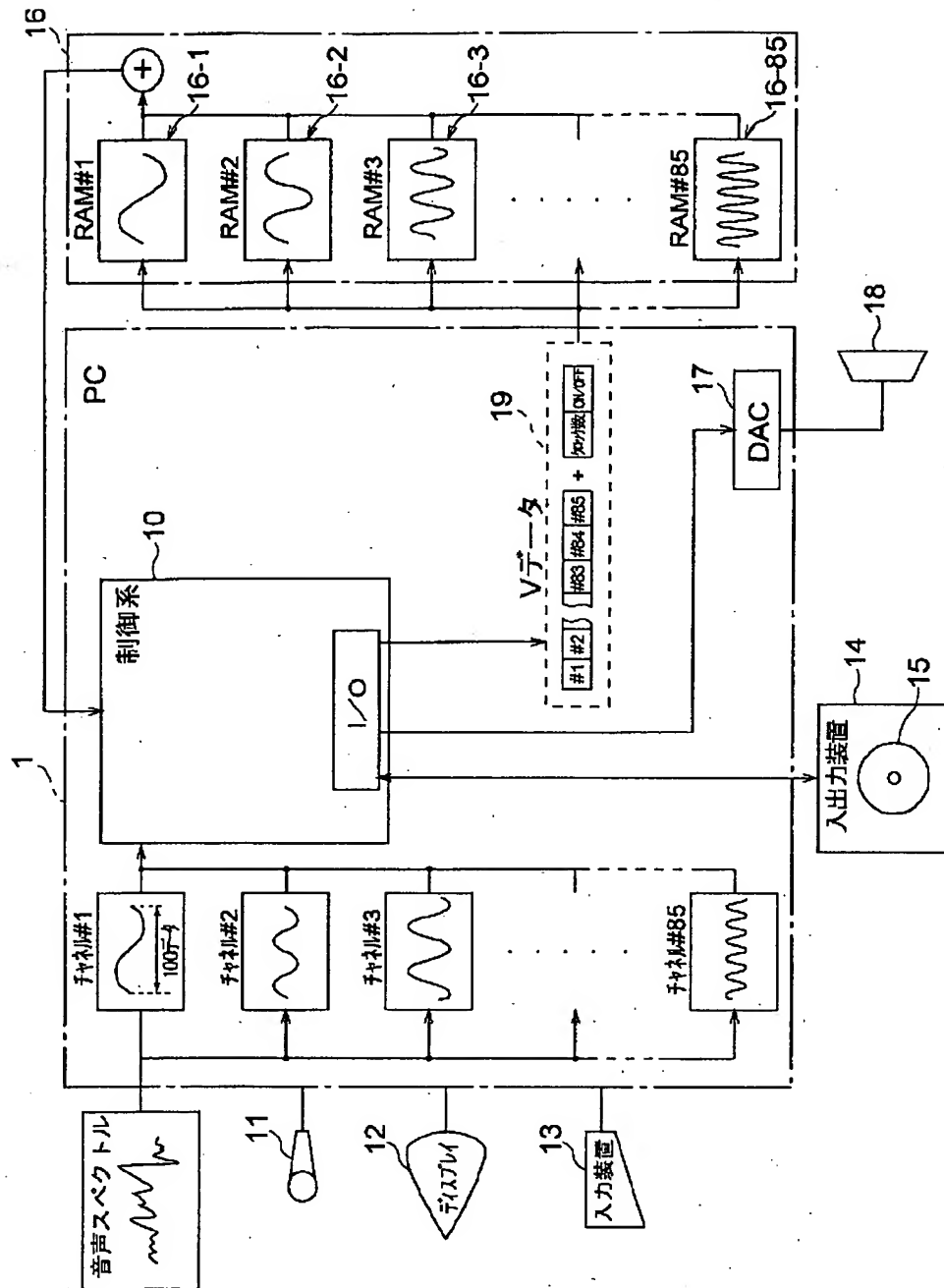
【図13】図12に示された再生装置の構成を示すブロック図である。

【図14】図13に示されたROM及びRAMのメモリ割り当て状況を説明するための図である。

【符号の説明】

1…PC、10…制御系、14…入出力装置、15…音声情報記録媒体、19…Vデータ、16-1~16-85…正弦波データ生成回路、17、175…DAC、18、177…スピーカ、173…加算器。

【図1】

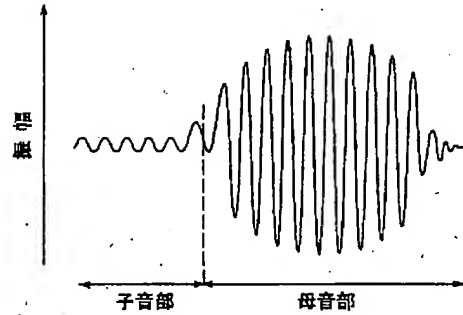




【図2】

チャンネル	低域f	高域f	中心f	音階	加減波 (44.1kHz)	加減数	波数	誤差 (ppm)
#1	75.57	80.06	77.78	D <sup>#</sup>	566.98	567	1	28.6
#2	80.06	84.82	82.41	E	535.13	535	1	242
#3	84.82	89.87	87.31	F	505.10	505	1	192
#4	89.87	95.21	92.50	F <sup>#</sup>	476.75	477	1	510
#5	95.21	100.9	98.00	G	450.00	450	1	0
#6	100.9	106.8	103.8	G <sup>#</sup>	424.88	425	1	340
#7	106.8	113.2	110.0					

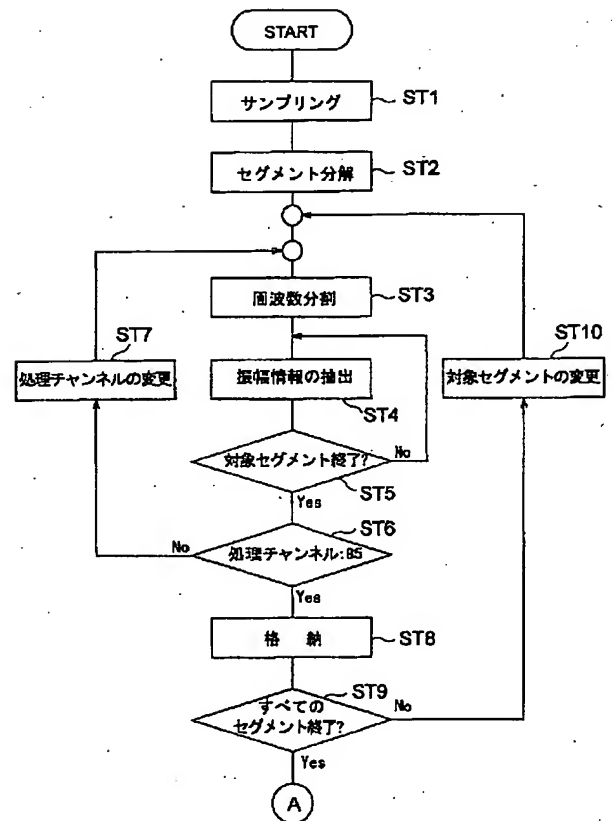
【図3】



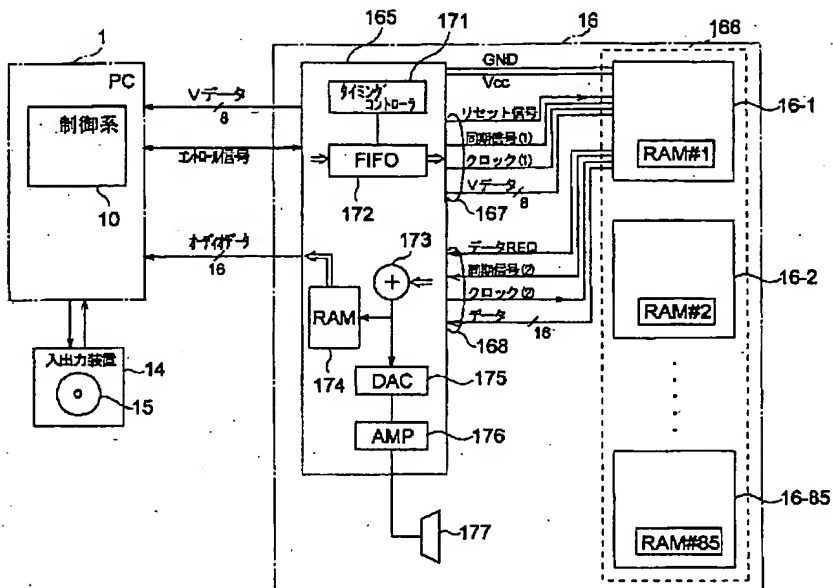
【図4】

- セグメント 621
- (a) It's not much of a problem.
- セグメント 622
- (b) I'd second that.
- セグメント 701
- (c) Yes!
- セグメント 799
- (d) I'm gonna go fishing with my brother.
- セグメント 801
- (e) I'm gonna go fishing with my brother,
- セグメント 802
- because he'd bought a new rod yesterday.

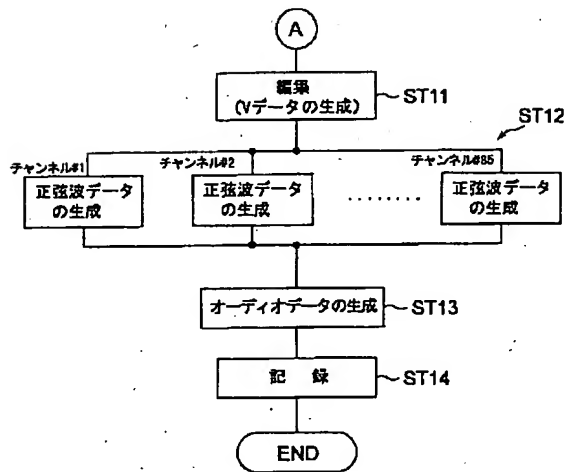
【図6】



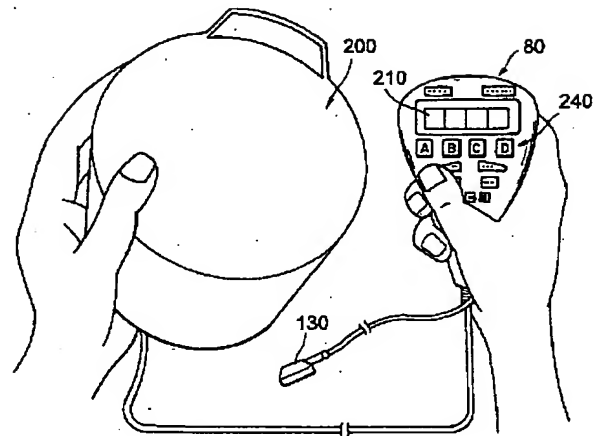
【図5】



【図7】



【図12】



【図9】

セグメント			621	622	
時間 (秒)	A		2.0	1.3	
	B		3.9	2.6	
	C		12	25	
容量 (KB)	A		12	7.8	
	B		23.4	15.6	
	C		72	150	

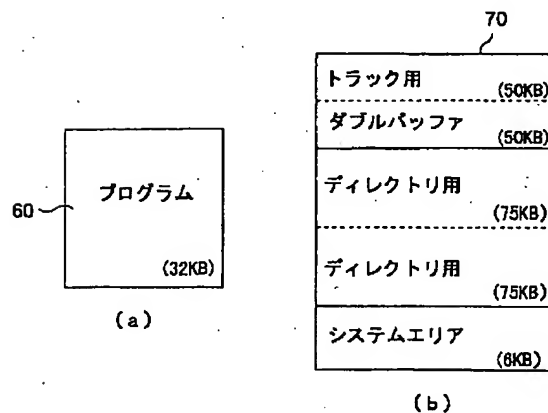
【図 8】

記録内容	
A	<div> <div>"It's not much of a problem."</div> <div>I'd second that."</div> </div> <div> <div>621</div> <div>622</div> </div>
B	<div> <div>It's...not...much...of...a...problem...</div> <div>I'd...second...that...</div> </div> <div> <div>621</div> <div>622</div> </div>
C	<p>(621) 「It」とは前の〇〇を指しており、「not much a problem」は慣用句で『問題がない』とか『心配ない』というような意味です。</p> <p>(622) 次に、「I'd」とは「I would」をつめた言い方であり、同じように「I could」も「I'd」とつめて言えます。 「second」とは『支持する』『賛成する』という意味があり、「I'd second that」で『同意です』という意味になります。」</p>

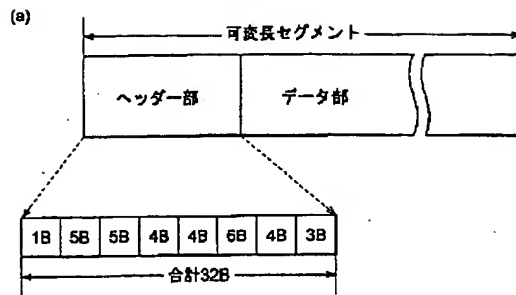
【図 10】

情報列	セグメント	カテゴリー	位置情報			スタートバイト	セグメント量 (Byte)
	SS	C	M	S	B	SB	LLL
A							
A	621	0	0	11	3	826	6,000
B		64	0	11	3	2,026	17,400
C		128	0	11	6	1,282	72,000
A	622	0	0	11	49	2,018	4,800
B		64	0	11	50	570	12,600
C		128	0	11	52	1,274	150,000

【図 14】



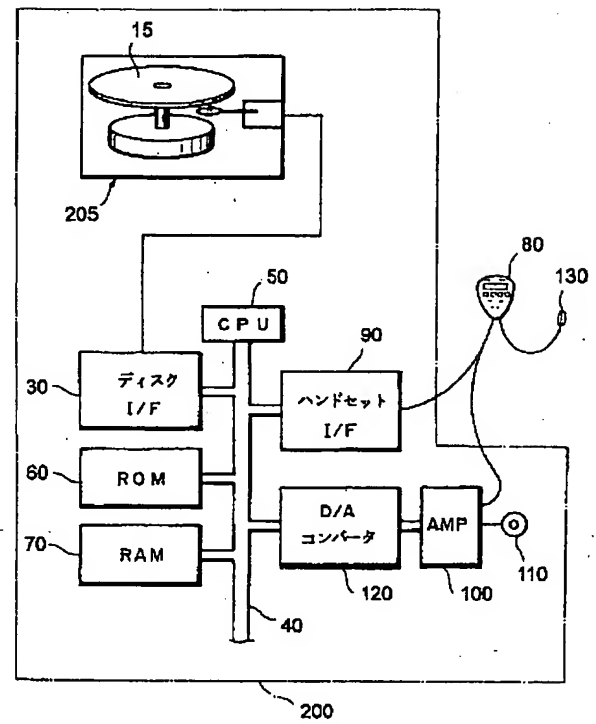
【図11】



(b)

ヘッダー (合計32B)	情報列	内 容
1B		文字情報や西暦情報の有無等 (1B)
5B	A	情報列タイプ (1B), データ長 (3B), 予備 (1B)
5B	B	情報列タイプ (1B), データ長 (3B), 予備 (1B)
4B	C	情報列タイプ (1B), データ長 (3B)
4B	(D)	情報列タイプ (1B), データ長 (3B)
6B	D	アドレス (3B), データ長 (3B)
4B	E	
3B	予備	

【図13】



*Date: December 26, 2003*

*Declaration*

*I, Michihiko Matsuba, President of Fukuyama Sangyo Honyaku Center, Ltd., of 16-3, 2-chome, Nogami-cho, Fukuyama, Japan, do solemnly and sincerely declare that I understand well both the Japanese and English languages and that the attached document in English is a full and faithful translation, of the copy of Japanese Unexamined Patent No. 2000-81897 laid open on March 21, 2000.*

A handwritten signature in cursive script, appearing to read 'm. matsuba', written in dark ink.

*Michihiko Matsuba*

*Fukuyama Sangyo Honyaku Center, Ltd.*

SPEECH SOUND DATA RECORDING METHOD, SPEECH SOUND DATA RECORDING  
MEDIUM, AND SPEECH SOUND DATA REPRODUCING METHOD AND UNIT  
Japanese Unexamined Patent No. 2000-81897

Laid-open on: March 21, 2000

Application No. Hei-10-249672

Filed on: September 3, 1998

Inventor: Hiroshi SEKIGUCHI

Applicant: Kanars Data Corporation

Patent Attorney: Yoshiki HASEGAWA et al.

#### SPECIFICATION

[TITLE OF THE INVENTION] Speech Sound Data Recording Method,  
Speech Sound Data Recording Medium, and Speech Sound Data  
Reproducing Method and Unit

#### [ABSTRACT]

[OBJECT] To provide speech sound data which has been  
emphasized/attenuated and has an extended/contracted  
reproduction period without changes in frequency components  
in an arbitrary part of the original speech sound data itself  
as speech sound for listening exercise of Japanese English  
learners.



[Solution Means] In a recording method relating to this invention, a first speech sound data sequence sampled in a first period is divided into a plurality of frequency components, and with respect to each frequency component, sine wave data which has been changed in the amplitude and wave number at a predetermined part is generated for an amplitude data sequence sequentially extracted from a second period, and a second speech sound data sequence synthesized by summing sine wave data corresponding to these frequency components is recorded on a predetermined recording medium. Thereby, the speech sound data sequence which has been extended/contracted and emphasized/attenuated at an arbitrary part is recorded on the obtained recording medium without changing the frequency.

[WHAT IS CLAIMED IS:]

[Claim 1] A speech sound data recording method for recording speech sound data on a predetermined recording medium, wherein a first speech sound data sequence sampled in a first period is divided into a plurality of frequency components, with respect to the plurality of frequency components, a corrected amplitude data sequence obtained by selectively editing one, two or more predetermined parts of an amplitude data sequence composed of amplitude data extracted in a second

period corresponding to at least one or more waveforms is generated,

V data composed of data component groups composed of amplitude data pieces extracted at the same timing corresponding to each other among the frequency components of the corrected amplitude data sequences of each of the plurality of frequency components and control data for instructing extension or contraction of the speech sound reproduction period on the basis of the first period, prepared for each data component group, is generated, sine wave data which has an amplitude given by the generated V data, has data intervals of the first period for each of the plurality of frequency components, and has a wave number corresponding to the reproduction period instructed by the control data included in the V data, is generated, and a second speech sound data sequence of the first period obtained by successively summing sine wave data generated for each of the plurality of frequency components is recorded on a predetermined recording medium.

[Claim 2] The speech sound data recording method according to Claim 1, wherein by changing amplitude values provided by amplitude data of selected parts corresponding to each other among the amplitude data sequences of the plurality of frequency components, a corrected amplitude data sequence is

generated for each of the plurality of frequency components.

[Claim 3] The speech sound data recording method according to Claim 1 or 2, wherein, each amplitude of sine wave data to be generated for each of the plurality of frequency components, is determined by a value obtained through linear interpolation between amplitude data pieces adjacent to each other of the corrected amplitude data sequence.

[Claim 4] The speech sound data recording method according to any one of Claims 1 through 3, wherein the control data includes frequency shift instruction data for reproducing the entirety of the plurality of frequency components by shifting them to higher tones or lower tones.

[Claim 5] The speech sound data recording method according to any of Claims 1 through 4, wherein the first speech sound data sequence corresponds to one, two or more sentences composed of word strings to be reproduced and outputted by a predetermined sound reproducing means, and is recorded on the predetermined recording medium in a condition where it is divided into variable-length sections for each of the data divided by vocalization pauses.

[Claim 6] The speech sound data recording method according to Claim 5, wherein the second speech sound data sequence is recorded on the predetermined recording medium for each section

divided according to the sections of the first speech sound data sequence, and this recording medium records recorded position identification data indicating switchable sections by means of recorded positions of these sections in this predetermined recording medium so that the first speech sound data sequence and the second speech sound data sequence are switched and reproduced by the predetermined sound reproducing means.

[Claim 7] A speech sound data recording medium on which a second speech sound data sequence has been recorded by the speech sound data recording method according to Claim 1.

[Claim 8] A speech sound data recording medium on which at least a second speech sound data sequence has been recorded whose amplitude and reproduction period have been selectively edited by changing at least the amplitude or waveform number in one, two or more parts corresponding to each other among frequency components composing a first speech sound data sequence sampled in a predetermined period.

[Claim 9] The speech sound data recording medium according to Claim 7 or 8, wherein the first speech sound data sequence corresponds to one, two or more sentences composed of word strings to be reproduced and outputted by a predetermined speech sound reproducing means, and has been recorded in a

condition where it is divided into variable-length sections for each of the data divided by vocalization pauses.

[Claim 10] The speech sound data recording medium according to Claim 9, wherein the second speech sound data sequence has been recorded for each section divided according to sections of the first speech sound data sequence, and recorded position identification data indicating switchable sections by means of recorded positions of the respective sections in this predetermined recording medium so that the first speech sound data sequence and the second speech sound data sequence are switched and reproduced by the predetermined sound reproducing means.

[Claim 11] A speech sound data reproducing method for reproducing a speech sound data sequence recorded in advance on a predetermined recording medium, wherein the recording medium contains: a first speech sound data sequence which corresponds to one, two or more sentences composed of word strings to be reproduced and outputted by a predetermined sound reproducing means, and is divided into variable-length sections for each of the data divided by vocalization pauses; a second speech sound data sequence which is divided according to the sections of the first speech sound data sequence, and has selectively changed amplitude strength

and reproduction period by changing at least the amplitude or waveform number of one, two or more parts corresponding to each other among frequency components composing the first speech sound data sequence; and recorded position identification data indicating switchable sections by means of recorded positions of these sections in this predetermined recording medium so that the first speech sound data sequence and the second speech sound data sequence are switched and reproduced by the predetermined speech sound reproducing means, and the speech sound data reproducing method comprises:

a first step in which, in response to a reproduction instruction for the second speech sound data sequence inputted during reproduction or after interruption of reproduction of the first speech sound data sequence, speech sound data sequences of sections in the second speech sound data sequence corresponding to the sections in the first speech sound data sequence that is being reproduced are readout from the recording medium based on the recorded position identification data, and the readout speech sound data sequences are reproduced by the predetermined sound reproducing means; and

a second step in which, in response to a reproduction instruction for the first speech sound data sequence inputted during reproduction or after interruption of reproduction of



the second speech sound data sequence, speech sound data sequences of sections in the first speech sound data sequence corresponding to sections of the second speech sound data sequence that is being reproduced are readout from the recording medium based on the recorded position identification data, and the readout speech sound data sequences are reproduced by the predetermined speech sound reproducing means.

[Claim 12] A speech sound data sequence reproducing unit for carrying out the speech sound data reproducing method according to Claim 11.

[DETAILED DESCRIPTION OF THE INVENTION]

[0001]

[Field of the Art] The present invention relates to a speech sound data recording method for recording speech sound data on a disk-type recording medium such as a CD-ROM, an MD, and an MO, etc., or a tape-type recording medium such as a DAT, etc., a speech sound data recording medium on which this speech sound data has been recorded, and a speech sound data reproducing method and unit for reading-out and reproducing speech sound data sequences recorded on the speech sound data recording medium.

[0002] Conventionally, for self-study of languages such as

English conversation, etc., exercise of recitation of a Chinese poem, self-study of laws, and other purposes, various teaching materials including speech sound data recorded on a recording medium such as a cassette tape, etc., have been provided. Herein, for example, for teaching materials for self-study of English conversation, a conventional major recording medium is, for example, a cassette tape (or record) recording a series of English vocal sound (speech sound data), and a learner has used this tape material in combination with a text. Such teaching materials on various levels from novice learners to expert learners are available.

[0003] Furthermore, Japanese Patent No. 2581700 proposes a data recording medium such as a CD-ROM, etc., comprising at least: a first region in which a speech sound data sequence (speech sound at a natural speed) that is divided into a plurality of sections and is suitable for learning of expert learners has been recorded; a second region in which a speech sound data sequence that is composed of equivalent sections corresponding to the abovementioned sections and suitable for learning of novice learners (including speech sound that is clear and has derivation differences although they have the same linguistic meanings); and a third region in which data has been recorded which indicates the relationship between the corresponding

sections of the abovementioned speech sound data sequences for learning of expert learners and novice learners by means of the recorded positions of the sections of these speech sound data sequences on the recording medium, and a reproducing method including switching reproduction between corresponding sections of the data recording medium having this structure.

[0004]

[Problem to be Solved by the Invention] As mentioned above, in the data recording medium of Japanese Patent No. 2581700, speech sound spoken by a native speaker have been recorded on the first region of the medium, and a speech sound data sequence composed of speech sound that has been slowed without changes in linguistic meaning has been recorded on the second region. Therefore, in a case where the speech sound cannot be heard during reproduction of the speech sound data sequence recorded on the first region, by reproducing it by switching it to the sound speech data sequence with the same content recorded on the second region (correspondence between the section that is being reproduced in the first speech sound data sequence and the section to be reproduced in the second speech sound data sequence has been recorded on the third region), a learner can understand the meaning of the speech sound that he/she could not catch.

[0005] However, although an English learner can understand the data recorded on the first region by listening to the data recorded on the second region as mentioned above, he/she is still not able to catch the sound of the data recorded on the first region, in particular, sound that the learner could not catch by only repeatedly listening to it. In a case of a Japanese English learner, it is generally known that a Japanese English learner is no good at catching phonemes that are not included in Japanese, in particular, consonants, and this obstructs conversations with a native speaker.

[0006] The present invention relates to a technique that improves listening comprehension of a learner to original speech sound that is difficult for the learner to catch by making the learner heard speech sound that has been edited in advance so as to become easier for the learner to catch, and an object of the invention is to provide a speech sound data recording method, a speech sound data recording medium, and a speech sound data reproducing method and unit for speech sound, as speech sound data for listening exercise of English learners, in which the amplitudes of frequency components and the reproduction period have been selectively edited without changing the frequency components of the original speech sound data itself.

[0007]

[Means for Solving the Problem] The present invention relates to a technique for newly generating, recording, and reproducing a speech sound data sequence in which a desired part in this speech sound data sequence has been emphasized or attenuated, and the reproduction period has been partially extended or contracted without changing the frequency components of the taken speech sound data sequence as sound for listening exercise. In this invention, since the sound quality of the speech sound data to be reproduced is not changed, desired editing is applied not to sampled speech sound data but to the respective frequency components of this speech sound data, and these edited frequency components are synthesized to obtain a new speech sound data sequence. By this construction, it becomes possible to provide speech sound data for listening exercise in which parts that are difficult for Japanese English learners to catch are selectively emphasized and/or extended. Furthermore, in a case where an expert learner wants to further improve his/her listening comprehension, on the contrary, it becomes possible to provide speech sound data whose sound has been selectively attenuated and reproduction period has been contracted.

[0008] Concretely, in a speech sound data recording method

relating to this invention, a first speech sound data sequence sampled in a first period (for example, 44.1kHz of audio clock of a music CD) is divided into a plurality of frequency components (hereinafter, referred to as channels), and in a second period (for example, corresponding to a necessary number of pieces of data for forming one waveform), amplitude data is obtained for each channel. This amplitude data is given in a form of an amplitude change amount of a waveform corresponding to, for example, 100 pieces of data of the first speech sound data sequence, and if one waveform is not formed by the 100 pieces of data, the number of pieces of data is increased so as to form one waveform (the second period is lengthened), and then the amplitude data is extracted. This second period is a patterned period.

[0009] Furthermore, a plurality of corrected amplitude data sequences are generated which have been edited so that the amplitude data is selectively changed for each of the amplitude data sequences of the respective channels thus obtained (amplitude data sequences extracted in the second period for each channel). The plurality of corrected amplitude data sequences are determined for each channel corresponding to each frequency component. Then, V data is generated which is composed of each data component group composed of amplitude



data, corresponding to each other among the corrected amplitude data sequences corresponding to the respective channels and extracted at the same timing, and control data, prepared for each data component group for instructing extension or contraction of the speech sound reproduction period on the basis of the first period.

[0010] Next, from the V data generated as data of the abovementioned second period, sine wave data are generated which have amplitudes (corrected amplitudes) given by the V data and data intervals of the first period, correspond to each channel, and have a wave number corresponding to a reproduction period instructed by the control data. The sine wave data thus generated for each channel are successively summed, whereby audio data of the first period (second speech sound data sequence) is generated. This generated audio data is recorded on a predetermined recording medium.

[0011] Furthermore, in the speech sound data recording method relating to this invention, each of the amplitude data in the amplitude data sequences of the respective channels extracted in the second period is edited so that an arbitrary part is selectively emphasized or attenuated. Namely, the speech sound data recording method relating to this invention generates corrected amplitude data sequences of amplitude data sequences

for the respective channels by selectively resetting amplitude values supplied from amplitude data of the predetermined parts corresponding to each other among the channels to be higher or lower. Furthermore, the speech sound data recording method relating to this invention is characterized in that, in order to avoid unnatural amplitude changes of reproduced sound, each amplitude of sine wave data to be generated for each channel is determined by a value obtained by linear interpolation between adjacent pieces of amplitude data of the corrected amplitude data sequences.

[0012] As mentioned above, the speech sound data recording method relating to the invention is constructed so that the amplitude of an arbitrary part of an amplitude data sequences generated for each channel is changed, and control data for instructing extension or contraction of the reproduction period is prepared for each of the data component groups collectively containing the amplitude data of each channel extracted in the second period, so that it becomes possible to selectively emphasize or attenuate the reproduced sound at an arbitrary part, and also to partially extend or contract the reproduction period without changing the frequency components.

[0013] The reason for this is that even when English at a natural

speed is only slowed and reproduced so that it can be heard by a Japanese, simple and uniform extension or contraction of the speech sound reproduction period of each frequency component is not sufficient, and depending on the type of vocal sound, a spectral time change at a consonant part results in a different sound linguistically. For example, comparing vocal sounds of "BA" and "PA," a spectral change of the former sound is quick, and the latter is only slow and its spectral form itself has almost no change. Therefore, when the vocal sound period of "BA" including the consonant part is extended, the sound sounds like "PA." In order to prevent this, the degree of extension of the consonant part is restricted to a limit at which the sound sounds "BA," and only the vowel part is extended or contracted to a desired sound reproduction period, whereby it sounds "BA." On the other hand, since the vowel part sounds as it is no matter how it is extended or contracted, the vowel part sound period can be set to a desired length (desired reproduction period). On the other hand, it is also necessary that only sound of a consonant part that is too weak for a Japanese to catch is selectively emphasized to the double or triple. Emphasizing sound including a vowel part is not effective since the entirety is excessively emphasized. Therefore, selective emphasizing is necessary. From the

abovementioned reasons, the speech sound data reproducing method relating to the invention is constructed so that for the amplitude data sequence of each channel, a corrected amplitude data sequence in which a part that is especially difficult for novice learners to catch is selectively emphasized is edited, and V data composed of amplitude data pieces generated at the same timing among the corrected amplitude data sequences for the respective channels and control data for instructing extension of the reproduction period are successively recorded. On the contrary, for expert learners, considering the properties of the respective vocal sounds, the speech sound data sequence may be selectively edited so that the reproduced speech sound is attenuated or the reproduction period is contracted at a desired part.

[0014] Furthermore, in the speech sound data recording method relating to this invention, in a case where a male voice is recorded on a predetermined recording medium according to the abovementioned recording method, if it is reproduced while extending the sound reproduction period, although the frequency spectrum of the outputted sound is invariable, a miscomprehension that the sound shifts to a lower sound may be sensuously caused. Therefore, it is preferable that the control data contains frequency shift instruction data for

enabling reproduction while shifting the entire frequency components to half- or one-higher or lower tones.

[0015] Furthermore, the technique relating to this invention is expected to have a dramatic learning effect in combination with the technique disclosed in the abovementioned Japanese Patent No. 2581700. By separately preparing speech sound data whose sound at an arbitrary part has been extended and/or emphasized according to variable-length sections divided by vocalization pauses spoken by a native speaker, it becomes possible for a learner to repeatedly reproduce and listen to sound that he/she could not catch, and by listening to speech sound which has been emphasized/extended at a part difficult to catch, improvement in listening comprehension for original sound is expected. Furthermore, for an expert learner, for more positive improvement in learning effect, by separately preparing speech sound data which has been contracted and/or attenuated at an arbitrary part in addition to vocal sound that was spoken by a native speaker and divided into sections, it becomes possible to reproduce sound by contracting the reproduction period and make it difficult to hear a consonant part (reduce the amplitude), and leaning in combination with vocal sound spoken by a native speaker becomes possible.

[0016] Concretely, the first speech sound data sequence

corresponds to one, two or more sentences composed of word strings to be reproduced and outputted by a predetermined speech sound reproducing means, and is recorded on the recording medium in a condition where it is divided into variable-length sections for each data divided by vocalization pauses. Thereby, the second speech sound data sequence is recorded on the predetermined recording medium for each of sections divided according to the sections of the first speech sound data sequence, and furthermore, this recording medium records recorded position identification data indicating switchable sections by means of recorded positions of these sections in the predetermined recording medium so that the first speech sound data sequence and the second speech sound data sequence are switched and reproduced by the predetermined speech sound reproducing means. By thus recording in advance the correspondence between the sections divided in the first speech sound data sequence and the sections divided in the second speech sound data sequence, one, two or more desired sections can be repeatedly reproduced, and also, it becomes possible to reproduce sound while switching a natural-speed reproducing sound and identical vocal sounds prepared for each level of learners in real time.

[0017] Therefore, by the recording method of the invention,

a speech sound data recording medium on which predetermined speech sound data (not waveform data but an amplitude data sequence with corrected frequency components) has been recorded is obtained.

[0018] As such a speech sound data recording medium, for example, a disk-type recording medium such as a CD-ROM, an MD, or an MO, etc., and a tape-type recording medium such as a DAT is applicable, and necessarily, on a related speech sound data recording medium, at least a second speech sound data sequence has been recorded which is synthesized by dividing the first speech sound data sequence sampled in a first period into a plurality of frequency components, generating sine wave data in which an amplitude at a predetermined part and a waveform number at a predetermined part are changed in amplitude data sequences successively extracted in a second period, and summing sine wave data corresponding to each of the frequency components. Namely, the second speech sound data sequence to be recorded on this speech sound recording medium has an amplitude and a reproduction period which have been selectively edited by changing at least the amplitude or changing the waveform number of parts corresponding to each other among frequency components composing a first speech sound data sequence sampled in a predetermined period.

[0019] Furthermore, this speech sound data recording medium relating to the invention becomes possible to be combined with the technique disclosed in the abovementioned Japanese Patent No. 2581700 by recording the abovementioned first data sequence corresponding to one, two or more sentences composed of word strings to be reproduced and outputted by a predetermined speech sound reproducing means in a manner in that the first data sequence is divided into variable-length sections for each data divided by vocalization pauses.

[0020] On the speech sound data recording medium having the abovementioned construction, in addition to the first speech sound data sequence, the abovementioned second speech sound data sequence has been recorded for each of sections divided according to the sections of the first speech sound data sequence, and furthermore, recorded position identification data has been recorded which indicates the switchable sections by means of recorded positions of these sections in this predetermined recording medium so that the first speech sound data sequence and the second speech sound data sequence are switched and reproduced by the predetermined speech sound reproducing means. Therefore, by preparing such a speech sound data recording medium, even during reproduction of one speech sound data sequence, a speech sound data reproducing method



and unit relating to the invention can switch it into a corresponding section of another speech sound data sequence and reproduce it in real time.

[0021] Furthermore, as an embodiment of the invention mentioned above, sales of recording software (a program which enables a personal computer to execute the abovementioned recording method, or a recording medium containing this program recorded), an exclusive recording unit, and a user's manual or a sale of a combination of these, a singular sale of a related speech sound data recording medium, and in addition, sales of this speech sound data recording medium, sound data reproduction software (including a program that can be executed by a personal computer or a recording medium containing this program recorded), an exclusive reproducing unit, and an user's manual, or a sale of a combination of these are considerable.

[0022]

[Embodiment of the Invention] Hereinafter, an embodiment of the invention is described with reference to Fig. 1 through Figs. 14. In the figures, the same components are attached with the same symbols and overlapping description is omitted.

[0023] The present invention relates to a technique which makes it possible to provide speech sound data in which a part that is hardly heard has been selectively emphasized or attenuated,

or the reproduction period thereof has been extended or contracted for, for example, listening exercise of English learners. Therefore, learners who heard speech sound data thus edited in advance can expect improvement in their listening comprehension to the original speech sound.

[0024] Fig. 1 is a block diagram for schematically describing speech sound data recording operation relating to the invention. First, speech sound (first speech sound data) at a natural speed spoken by a native speaker sampled in, for example, 44.1kHz (first period) of an audio clock of a music CD is taken into the main body of a PC 1, and temporarily recorded on its hard disk, etc. Then, the taken-in speech sound data is filtered so as to be divided into channels (frequency components) as shown in the table of Fig. 2. The range of frequency of speech sound data to be taken-in is between 75Hz and 10,000Hz, and the sampling frequency is set to 44.1kHz (22.68 $\mu$ s) in accordance with the audio clock of a music CD. The number of divided channels is set to 85 (7 octaves and 1 tone), and the central frequencies (central f) of the respective channels #1 through #85 are set to be in a halftone line (77.78Hz(D#) to 9,960Hz(D#)) of equal temperament (equal temperament of 12 degrees per 1 octave).

[0025] As mentioned above, as for the data divided into the

channels #1 through #85, amplitude data thereof is extracted every 2.268ms (corresponding to 100 pieces of data in 44.1kHz sampling, however, the number of pieces of data is increased when 100 pieces of data cannot form one waveform). Therefore, in this embodiment, the sampling rate (second period) of the amplitude data in each of the channels #1 through #85 is 441 samples/s (2.268ms). This sampling rate is a patterned period, and the embodiment can repeat processing alternately at difference rates, for example, after 100 pieces of data are taken-in, 120 pieces of data are taken-in and processed.

[0026] Furthermore, a control system 10 of the PC 1 applies various editing (can be carried out via a display 12 and an input device 13 such as a key board and a mouse, etc.) to the amplitude data of each of the channels #1 through #8 sampled every 2.268ms to generate corrected amplitude data groups at every 2.268ms. Then, the corrected amplitude data (elements composing the corrected amplitude data groups) for each of the channels #1 through #85 are expressed by 1 byte (8 bits), respectively, and control data with 2 bytes is further added to generate V data 19 with 87 bytes ( $85 \text{ channels} \times 1 \text{ byte} + 2 \text{ bytes}$ ).

[0027] Furthermore, the corrected amplitude data is obtained by editing each amplitude data in the amplitude data sequences

of the channels #1 through #85 (amplitude data sampled every 2.268ms) so that arbitrary parts of them are selectively emphasized or attenuated. Namely, the amplitude data sequences of the channels #1 through #85 are generated by selectively resetting amplitude values given from the amplitude data of predetermined parts corresponding to each other among the channels to be higher or lower. Furthermore, the abovementioned control data is composed of extension instruction data (1 byte) for instructing extension or contraction of each reproduction period of the frequency components of the channels #1 through #85 and frequency shift instruction data (1 byte) for instructing whether or not the frequency components corresponding to the respective channels #1 through #85 are entirely shifted to half- or one-lower or higher tones and reproduced.

[0028] The abovementioned extension instruction data is expressed by the number of reproduction clocks that indicates how many milliseconds one piece of data is to be reproduced in. For example, when this extension instruction data is expressed by a half of the number of clocks to be reproduced, the reproduction period becomes equivalent to the original period when the data is 50, and when this data is set to 100, reproduction is carried out in 200 clocks of 44.1kHz and the

reproduction period can be extended to double (this data is expressed by 1 byte, so that extension up to  $256/50 = 5.12$  times is possible at a maximum). On the contrary, when this data is set to 25, reproduction is carried out in 50 clocks of 44.1kHz, whereby the reproduction period can be contracted to the half. Furthermore, the frequency shift instruction data is set as ON "1" when the whole frequency components are shifted to lower or higher tones, and is set as OFF "0" when shifting is not necessary.

[0029] As mentioned above, based on V data 19 generated by applying desired editing by the control system 10 to the amplitude data of the channels #1 through #85 sampled every 2.268ms, a new speech sound data sequence is generated.

[0030] In order to generate a new speech sound data sequence from the generated V data 19, an external device 16 having sine wave generating circuits 16-1 through 16-85 for generating sine waves with wavelengths corresponding to the respective channels #1 through #85 becomes necessary. The generating circuits 16-1 through 16-85 have ROMs on which basic data on sine waves with frequencies corresponding to the respective channels #1 through #85 have been recorded and RAMs #1 through #85 for temporarily recording generated sine wave data, respectively, and in these circuits, sine wave data formed

based on corrected amplitude data of the V data 19 supplied from the control system 10 and has a waveform number instructed by extension instruction data of control data are written on the respective RAMs #1 through #85. Intervals of the data composing this sine wave data are 22.68 $\mu$ s that is the data interval of the sampling frequency 44.1kHz.

[0031] Then, the sine wave data written on the RAMs #1 through #85 in these generating circuits 16-1 through 16-85 are successively readout at 44.1kHz timings, and are summed to generate audio data (speech sound data sequence). This audio data is transmitted to the control system 10, and outputted from the control system 10 to the input/output device 14 such as a CD-ROM writing device, etc., via an I/O. This I/O device 14 records the audio data of 44.1kHz transmitted from the control system 10 on a predetermined speech sound data recording medium 15 such as a CD-ROM, etc.

[0032] In generation of sine wave data at the respective generating circuits 16-1 through 16-85, in order to avoid unnatural amplitude changes in the reproduced sound, the amplitudes of the sine wave data of the respective channels #1 through #85 are determined by values obtained by linear interpolation between adjacent pieces of amplitude data of the corrected amplitude data sequence. Furthermore, the audio data

generated in the external device 16 may be directly outputted from the speaker 18 via a DAC 17 and an AMP in a form of speech sound. Furthermore, as such a speech sound data recording medium 15, for example, a disk-type recording medium such as a CD-ROM, an MD, an MO, etc., and a tape-type recording medium such as a DAT, etc., are applicable.

[0033] On the other hand, this invention mainly relates to a technique that simply enables slow reproduction of natural-speed English so that it can be heard by a Japanese, however, only simple and uniform extending or contracting the sound reproduction period for each frequency component is not sufficient. Namely, Fig. 3 shows the basic form of a speech sound spectrum, and depending on the type of vocal sound, a spectral time change at a consonant part results in a different sound linguistically. For example, comparing vocal sounds of "BA" and "PA," a spectral change of the former sound is quick, and the latter is only slow and its spectral form itself has almost no change. Therefore, when the vocal sound period of "BA" including the consonant part is extended, the sound sounds like "PA." In order to prevent this, the degree of extension of the consonant part is restricted to a limit at which the sound sounds "BA," and only the vowel part is extended or contracted to a desired speech sound reproduction period,

whereby it sounds "BA." On the other hand, since the vowel part sounds as it is no matter how it is extended or contracted, the vocal part sound period can be set to a desired length (desired reproduction period). On the other hand, it is also necessary that only sound of a consonant part that is too weak for a Japanese to catch is selectively emphasized to the double or triple (by increasing the amplitude). Emphasizing sound including a vowel part is not effective since the entirety is excessively emphasized. Therefore, selective emphasizing is necessary. From the abovementioned reasons, the speech sound data reproducing method relating to the invention is constructed so that for the amplitude data sequence of each channel, a corrected amplitude data sequence in which a part that is especially difficult for novice learners to catch is selectively emphasized is edited, and V data composed of amplitude data pieces generated at the same timing among the corrected amplitude data sequences for the respective channels and control data for instructing extension of the reproduction period are successively recorded. On the contrary, for expert learners, considering the properties of the respective vocal sounds, the speech sound data sequence may be selectively edited so that the reproduced speech sound is attenuated or the reproduction period is contracted at a desired part.



[0034] Furthermore, in this invention, in a case where a male voice is recorded on a predetermined recording medium according to the abovementioned recording method, if it is reproduced while extending the speech sound reproduction period and/or emphasizing speech sound at an arbitrary part, although the frequency spectrum of the outputted speech sound is invariable, a miscomprehension that the sound shifts to a lower sound may be sensuously caused. On the contrary, if reproduction is carried out while contracting the speech sound reproduction period and/or attenuating the speech sound at a desired part, a miscomprehension that the sound shifts to a lower sound may be sensuously caused. Therefore, the control data contains frequency shift instruction data for enabling reproduction while shifting the entire frequency components to half- or one-higher or lower tones.

[0035] Next, the present invention is suitable for the technique for reproduction, etc., of a recording medium on which speech sound spoken by a native speaker has been recorded as disclosed in the abovementioned Japanese Patent No. 2581700. Hereinafter, the construction for applying this invention to such a technique is described.

[0036] This invention can be expected to show a remarkable learning effect in conjunction with the technique disclosed

in the abovementioned Japanese Patent No. 2581700. Namely, by separately preparing speech sound data which has been selectively extended or contracted and emphasized or attenuated at an arbitrary part in accordance with variable-length sections divided by vocalization pauses of vocal sound spoken by a native speaker, not only a learner can repeatedly reproduce and listen to speech sound that he/she could not hear, but also can listen to reproduced speech sound that has been extended or contracted and emphasized or attenuated at a part that is difficult for the learner to catch and improve his/her listening comprehension.

[0037] Figs. 4 are for schematically describing various data including speech sound data sequences to be recorded on a speech sound data recording medium relating to the invention.

[0038] First, a first speech sound data sequence (speech sound data sequence sampled at 44.1kHz) to be recorded on the speech sound data recording medium 15 is a series of speech sound data composed of a plurality of sentences with different lengths such as conversations among casts of a film and conversations in usual life environment, and between the sentences (speakers' speech sound data), a voiceless period randomly occurring such as a condition without speech sound reproduced, a condition with only noise, a condition with only music (BGM) reproduced,

etc., can exist. Therefore, the first speech sound data sequence corresponds to one, two or more sentences composed of a plurality of word strings to be reproduced and outputted by a predetermined speech sound reproducing means, and is recorded on a first region of the speech sound data recording medium 15 in a condition where, as shown in Figs. 4, the sequence is divided into variable-length sections (hereinafter, referred to as segments) for each speech sound data divided by vocalization pauses.

[0039] Generally, in English conversation made by native speakers, since one sentence is usually vocalized in approximately three seconds, it is reasonable that vocalization pauses that determine the segments composing the speech sound data sequence to be recorded are set between sentences to compose the variable-length segments 621, 622, and 799 that composes the speech sound data sequence as shown in Fig. 1(a), Fig. 1(b), or Fig. 1(d). The sentences in conversation also include extremely short sentences as shown in Fig. 1(c), and this sentence 701 also composes one segment. On the other hand, as shown in Fig. 1(e), in a case of an extremely long sentence, since points before conjunctions and relatives become vocalization pauses, it is reasonable that the sentence shown in Fig. 1(e) is composed of successive two

segments 801 and 802. Therefore, segments of a speech sound data sequence to be recorded means speech sound data recording units divided based on vocalization breaks (breathing points) or certain linguistic (grammatical) breaks.

[0040] In the speech sound data recording method relating to the invention, first, as mentioned above, a second speech sound data sequence is generated which has been selectively edited (amplitude change and reproduction period change of each frequency component) at an arbitrary part with respect to each segment obtained by dividing the first data sequence as mentioned above. Concretely, this second speech sound data sequence is recorded on a predetermined speech sound data recording medium 15 by a device comprising the main body of a PC 1 that edits the respective frequency components and an external device 16 that generates edited audio data (second speech sound data sequence) as shown in Fig. 5.

[0041] Particularly, the external device 16 comprises, as shown in Fig. 5, a master board 165 for generating audio data and a slave board 166 including sine wave generating circuits 16-1 through 16-85 provided for each channel. The master board 165 has a timing controller 171 and an FIFO 172 for supplying the respective generating circuits 16-1 through 16-85 with V data from the PC 1 in response to control signals, and has an adder

173 for successively summing sine wave data (16 bits) transmitted from the respective generation circuits 16-1 through 16-85 to generate audio data (16 bits) and a RAM 174 as a buffer for temporarily storing the generated audio data to be transmitted to the PC 1. Furthermore, the master board 165 shown in Fig. 5 is provided with a DAC 175 and an AMP 176 to make a speaker 177 to directly output and reproduce the generated audio data so that a first speech sound data sequence and a newly edited second speech sound data sequence are outputted over any number of times as sound from the speaker, heard by ears and compared with each other (they may be provided at the PC 1 side in the structure for speech sound reproduction as shown in Fig. 1). On the other hand, the slave board 166 has the sine wave generating circuits 16-1 through 16-85 which generate sine waves with predetermined frequencies for each channel, and these generating circuits 16-1 through 16-85 have ROMs on which data for generating sine waves have been recorded and RAMs #1 through #85 as buffers for temporarily storing generated sine wave data, respectively.

[0042] Furthermore, the master board 165 and the slave board 166 are connected by a total of 32 buses including 30 signal buses and a GND and a Vcc, and the figures, the bus group shown by the reference numeral 167 is a V data related bus group for

supplying V data to the respective generating circuit 16-1 through 16-85, and the bus group shown by the reference numeral 168 is an audio data related bus group for transmitting sine wave data for audio data generation from the respective generating circuits to the master board 165.

[0043] Next, an embodiment of the speech sound data recording method relating to this invention, applied to the technique disclosed in Japanese Patent No. 2581700, is described by using the flowcharts of Fig. 6 and Fig. 7 with reference to Fig. 5.

[0044] First, generation of V data is carried out by the PC 1 side. Namely, in the PC 1, a series of speech sound data (a first speech sound data sequence) is sampled in 44.1kHz (16bit/data), and sampling data corresponding to this first speech sound data sequence is temporarily stored in the hard disk (Step ST1), and divided into a plurality of segments as shown in Figs. 4 (Step ST2).

[0045] Subsequently, as for one of the divided segments, according to a digital band pass filter program, waveform data of the band width (75.57kHz through 80.06kHz) of the first channel #1 is developed on a memory. At this point, development is carried out while maintaining the data intervals corresponding to the rate of 44.1kHz. Then, average amplitude data (8 bits) is extracted every 100 pieces of data (Step ST3).

In a case where for the frequency components of the first channel #1, 100 pieces of data cannot form one waveform, an amplitude data is obtained by increasing the number of pieces of data so as to form one waveform. Until the sampling data of the target segment is finished, the same operation is repeated while shifting 100 pieces of data. By this operation, amplitude data sequence with data intervals of 2.268ms (441 pieces of amplitude data per one second) for the first channel #1 that is the target channel is obtained. When amplitude data extraction for the first channel #1 that is the target channel is finished (Step ST5), subsequently, a frequency of the second channel #2 is divided by the digital band pass filter and the operations of Step ST3 through ST5 are repeated, and while changing the target channel (Step ST7), amplitude data sequences for target segments of the first channel #1 through the 85th channel are generated.

[0046] The abovementioned operation is carried out while the target segment is changed (Step ST10) until all segments composing the first speech sound data sequence sampled in Step ST1 are finished (Step ST9).

[0047] Next, editing described below is applied at the PC 1 side to the amplitude data sequences of 85 channels for the respective segments obtained by executing the abovementioned

Steps ST1 through ST9, whereby V data is generated (Step ST11).

[0048] First, an amplitude data sequence group for 85 channels, generated for each of the divided segments is called from the hard disk that has stored the group, and its amplitude waveforms are successively displayed on a monitor 12.

[0049] In actual editing, a desired part of the displayed amplitude waveform is specified and a reproduction period is specified (based on 50 clocks). Furthermore, if necessary, the amplitude is changed (set by a magnification based on the displayed amplitude waveform) by specifying a part to be changed, and an instruction of frequency shift to lower or higher tones is specified. The amplitude data sequence obtained for a selected arbitrary part is edited to generate new corrected amplitude data in which each amplitude data has been corrected, for example, for a consonant part of the segments, the amplitude is doubled and the reproduction period is increased by half, on the other hand, at a vowel part, the amplitude is not changed and only the reproduction period is increased to 2.5 times.

[0050] Then, in the obtained corrected amplitude data sequences for 85 channels, control data composed of data for instructing the abovementioned reproduction period change and data for instructing frequency shift are added to each data component



group collectively containing data components at the same timing among the corrected amplitude data sequences, whereby V data with data intervals of 2.268ms is obtained.

[0051] Next, the V data (87 bytes/data) thus prepared at the PC 1 side is transmitted to the master board 165 of the external device 16, and further transmitted from the master board 165 to the respective sine wave generating circuits 16-1 through 16-85 on the slave board 166 via the data buses. The slave board 166 comprises 11 boards each of which has 8 circuits mounted (eleventh board uses only 5 circuits of the 8 circuits), and each circuit generates sine wave data for a corresponding channel (Step ST12). The circuits have the same structure except that the ROMs storing wave form data of sine waves are different and the 7-bit DIP-SW settings specifying corresponding channels are different.

[0052] First, the circuits assigned for the respective channels commonly receive the header (2 bytes) of 87-byte V data supplied from the master board 165, and on the other hand, as for the corrected amplitude data of the V data, they receive only the corrected amplitude data (1 byte) for the corresponding channels. In each circuit, a reproduction period of the received header data is investigated in order to judge how many clocks of 44.1kHz outputting of the formation of a waveform

needs. For example, when the instructed reproduction period is given as 50, 100-clock reproduction is carried out (the reproduction period does not change), and in the case of 110, 220-clock reproduction (the reproduction period doubles) is carried out. In each circuit, sine wave data of a corresponding frequency has been recorded on the ROM at data intervals when outputted at 44.1kHz (M waves of the sine wave of this frequency are accurately stored from the address 0 to the address N of the ROM (M and N are natural numbers). The processor in each circuit pluses 1 to the address of the ROM every time it generates one piece of sine wave data (every 22.68 $\mu$ s). Then, next to the address N, it returns to 0. Thereby, accurate sine waves can be generated without discontinuous points. However, regarding the abovementioned sine wave data, one piece of sine wave data is generated by multiplying basic data stored in the ROM by the received corrected amplitude data. Furthermore, each of the amplitude data is a value obtained through linear interpolation between the current amplitude data and the previous amplitude data.

[0053] As mentioned above, the sine wave data generated in each circuit is stored in the respective RAMs #1 through #8 that are output buffers by referring to the ROM in the period of 44.1kHz (22.68 $\mu$ s) and multiplying referred data by a

coefficient obtained through the abovementioned interpolation.

[0054] Then, a control signal from the master board 165 gives a timing ( $22.68\mu\text{s}$  period) for outputting sine wave data stored in the respective RAMs #1 through #85 to the output buses (16 bits), and only at the timing, the data are transmitted to the buses. The time width given for one circuit is  $226\text{ns}$  ( $22.68\mu\text{s}/85$ ). On the other hand, the taking-in timings at the master board 165 side are given by clocks and synchronizing signals. The number of clocks from the synchronizing signal becomes the same as the channel number specified by the DIP switch. Two types of waveform data are stored in the ROM provided in each circuit so that sine wave data with a frequency shifted by a half tone (or one tone) can be outputted from each circuit when the frequency shift instruction data contained in the header data of the V data is ON, and either of the types can be selected.

[0055] On the other hand, the master board 165 receives sine wave data generated in each of the circuits 16-1 through 16-85 at a rate of 85 pieces of data in  $22.68\mu\text{s}$  (data interval:  $22.68\mu\text{s}/85 = 266\text{ns}$ ). In actuality, the master board successively sums sine wave data from the respective circuits while receiving them to generate audio data of  $44.1\text{kHz}$  (second

speech sound data sequence) (Step ST13). The generated audio data is successively stored in the RAM 174 that is a buffer, and then transmitted to the PC 1.

[0056] The PC 1 records the transmitted audio data on a predetermined recording medium 15 while controlling the I/O device 14 (Step ST14), whereby a speech sound data recording medium relating to the invention is obtained.

[0057] Next, examples in which the technique disclosed in the abovementioned Japanese Patent No. 2581700 is applied to the speech sound data recording medium of the invention are described.

[0058] First example relating to the speech sound data recording medium

First, in a first example, at least two types of speech sound data sequences and recorded position identification data are recorded. Namely, the first speech sound data sequence is composed of English speech sound data spoken by a native speaker at a natural speed, and this speech sound data sequence is divided into a plurality of variable-length segments by vocalization pauses (sentence ends or vocalization or grammatical breaks at which a speaker takes a rest during sentence) as mentioned above. A second speech sound data sequence is obtained by selectively editing an arbitrary part

of the first data sequence as mentioned above, and is divided into a plurality of variable-length segments according to the respective segments of the first speech sound data sequence. Furthermore, the recorded position identification data indicates at what positions the respective segments in the first and seconds speech sound sequences have been recorded on the speech sound data recording medium. Therefore, for example, the recorded position of the segment "It's... not... much... of... a... problem." of the second speech sound data sequence corresponding to the t-th segment "It's not much of a problem." of the first speech sound data sequence in the medium can be recognized by this recorded position identification data.

[0059] As a result, the first and second speech sound data sequences and the recorded position identification data are not recorded independently from each other, but are recorded in a fixed relationship, and the speech sound data sequences are organically combined by using the segments as units. Namely, the first and second speech sound data sequences are paired each other, and the recorded position identification data links these sequences to each other for each segment. In this example, the recorded position identification data is recorded on a directory region of the speech sound data recording medium, and contains at least data on the header positions of the

respective segments.

[0060] In the reproduction method for the speech sound data recording medium (first example) having the abovementioned structure, speech sound is reproduced in order by recorded segment, and in particular, this reproducing method is characterized in that reproduction switching from the first speech sound data sequence to the second speech sound data sequence recorded on the speech sound data recording medium (or reproduction switching from the second speech sound data sequence to the first speech sound data sequence) is possible. This reproduction switching is carried out by a per-segment basis. For example, when a reproduction instruction for the second speech sound data sequence is inputted during reproduction of the t-th segment of the first speech sound data sequence (arise of an interrupt request), the corresponding t-th segment of the second speech sound data sequence is readout based on the recorded position identification data, and speech sound reproduction of the corresponding segment is executed. On the contrary, reproduction switching from the second speech sound data sequence to the first speech sound data sequence is also carried out on a per-segment basis in the same manner as the abovementioned reproduction switching.

[0061] Furthermore, in this reproducing method, in addition

to the abovementioned reproduction switching, various changes into repeating reproduction, etc., are possible. As a typical example, a so-called return instruction is available. Namely, when a return instruction is inputted after interruption of the reproduction in response to a stop order during reproduction, by returning the speech sound data reading-out position by an instructed amount, speech sound data reproduction satisfying the request of an operator is carried out.

[0062] Second example relating to the speech sound data recording medium

This second example has a structure basically similar to that of the abovementioned first example, however, it is characterized by including, in addition to the first speech sound data sequence and the second speech sound data sequence, a third speech sound data sequence that is separate from the first speech sound data although having semantic contents equivalent to those of the first speech sound data, and is speech sound data of slow-speed English spoken by, for example, spacing words. This third speech sound data sequence is also composed of a plurality of variable-length segments, and the recorded position identification data controls recorded positions of the first through third speech sound data

sequences among these segments. Therefore, the speech sound data reproducing method in this second example is the same as in the first example.

[0063] Furthermore, in this example, it is important that the first speech sound data sequence and the third speech sound data sequence are divided into a plurality of variable-length segments, respectively, and semantic contents of the segments correspond to each other between these sequences. For example, when the t-th segment (621st segment in Fig. 4(a)) of the first speech sound data sequence is "It's not much of a problem." spoken by a native speaker, the t-th segment of the third speech sound data sequence is "It is not much of a problem." spoken by spacing the words. However, the fact that the third speech sound data sequence has contents corresponding to those of the second speech sound data sequence and is composed of different speech sound data indicates that it has the same linguistic meaning and is vocalized differently.

[0064] Third example relating to the speech sound data recording medium

Furthermore, a third example of the speech sound data recording medium relating to the invention is described. The speech sound data recording medium relating to this third example has a different point from the abovementioned speech sound data



recording medium of the first example in that a fourth speech sound data sequence containing grammatical explanation, etc., has been recorded on the speech sound data recording medium in addition to the first and second speech sound data sequences.

[0065] The important fact herein is that the third speech sound data sequence is divided into segment groups each of which collectively includes one, two or more variable-length segments of the first and second speech sound data sequences. In other words, one segment group of this fourth speech sound data includes one, two or more segments of the first and second speech sound data sequences, and therefore, one segment group of the fourth speech sound data sequence is paired with one, two or more segments of the first and second speech sound data sequences. Particularly, in this construction, as shown in Fig. 4(e), it is assumed that one sentence is divided into a plurality of segments.

[0066] Furthermore, in the speech sound data recording medium of this third example, the recorded position identification data recorded on a predetermined region also includes data indicating recorded positions of the contents of the fourth speech sound data sequence for each segment group. Therefore, the first, second, and fourth speech sound data sequences and the recorded position identification data are recorded on the

medium in a fixed relationship, and the speech sound data sequences are organically combined by a per-segment or per-segment group basis. In this third example, the recorded position identification data is also recorded on a directory region of the speech sound data recording medium, and includes data on the header positions of the segments in each speech sound data sequence. It is also possible that the third speech sound data that is equivalent to the speech sound data of the first speech sound data sequence and is made by speaking at a low speed while spacing the words is further recorded.

[0067] The reproducing method for the speech sound data recording medium (third example) having the abovementioned structure is basically the same as in the first example mentioned above except that in addition to reproduction switching between the first and second speech sound data sequences, reproduction switching is also carried out between the first/second speech sound data sequence and the fourth speech sound data sequence.

[0068] For example, in a case where "It's not much of a problem." spoken by a native speaker cannot be heard during reproduction of the first speech sound data sequence, by switching the reproduction from the first speech sound data sequence that is being reproduced currently to the second speech sound data

sequence, the speech sound "It's... not... much...of...a... problem." selectively subjected to editing such as extension can be listened to. Then, when a user wants to know the meaning in Japanese and the grammar of this sentence, he/she switches the reproduction from the speech sound data sequence currently reproduced to the fourth speech sound data sequence. As a matter of course, in this reproducing method, combined use of a return instruction and a stop order described in the speech sound data recording medium reproducing method of the first example mentioned above can be applied and used. Furthermore, in this reproducing method, switching reproduction and repeating reproduction are possible.

[0069] Fourth example relating to the speech sound data recording medium

The fourth example of the speech sound data recording medium relating to the invention is basically the same as the first example, however, its main difference from the first example is in that a text data sequence has been recorded in addition to the first and second speech sound data sequences. This text data sequence corresponds to text data whose contents correspond to the first or second speech sound data sequence, for example, text data corresponding to English (speech sound) spoken by a native speaker.

[0070] This text data sequence is also divided into segments corresponding to the segments of the first and second speech sound data sequences. In the speech sound data recording medium relating to this fourth example, recorded position identification data also includes data on header positions indicating recorded positions of the segments of this text data sequence corresponding to the segments of each speech sound data sequence, and is recorded on a directory region of the speech sound data recording medium. Therefore, the first and second speech sound data sequences and the text data sequence correspond to each other on a per-segment basis.

[0071] Furthermore, in the speech sound data recording medium relating to the fourth example, in a case where the fourth speech sound data sequence in the abovementioned third example is added as recording data, one, two or more segments of the first and second speech sound data sequences and the text data sequence also correspond to one segment group of the third speech sound data sequence. In this construction, the recorded position identification data includes the header positions of the respective segments, and is recorded on a directory region of the speech sound data recording medium. Then, in the same manner as in the third example mentioned above, a third speech sound data sequence that is equivalent to the speech sound data

of the first speech sound data sequence and is spoken at a low speed by spacing words can be recorded in this fourth example.

[0072] The reproducing method for the speech sound data recording medium (fourth example) having the abovementioned structure is also basically similar to in the case of the abovementioned second example, however, it has a difference in that the text data sequence is displayed during reproduction of the first or second speech sound data sequence.

[0073] For example, during reproduction of the segment "It's not much of a problem." of the first speech sound data sequence, "It's not much of a problem." or "It is not much of a problem." is displayed on a predetermined display part. This display does not have to be completely temporally synchronized with the speech sound data sequence that is being reproduced, and texts may be displayed by being successively delayed or advanced. In this reproducing method, switching reproduction and repeating reproduction are also possible.

[0074] Next, a detailed structure of the speech sound data recording medium relating to the invention is described in detail with reference to Fig. 8 through Figs. 11.

[0075] Fig. 8 shows an example of the speech sound data recording medium relating to the invention for explaining speech sound data sequences A, B, and C when the abovementioned

third example is applied to English conversation self-study and corresponding recorded contents. In this figure, the speech sound data sequence A is a data sequence (first speech sound data sequence) of English spoken by a native speaker, and is composed of a plurality of segments 621 and 622. The speech sound data sequence B is a data sequence (second speech sound data sequence) edited so that a predetermined part of the first data sequence is selectively extended as described above with reference to the flowcharts of Fig. 6 and Fig. 7. The speech sound data sequence C is a data sequence (third speech sound data sequence) for explaining in Japanese, and segment groups included in this speech sound data sequence C correspond to the segments 621 and 622 of each of the speech sound data sequences A and B, respectively.

[0076] Fig. 9 is a table for describing the relationship between time and capacity per segment in the case shown in Fig. 8. In this table, one second corresponds to a capacity of 6 kilobytes. For example, in the segment 621 of the speech sound data sequence A, the vocalization time of "It's" is 0.2 seconds and the capacity thereof is 1.2kB (kilobytes), the vocalization time of "not" is 0.1 seconds and the capacity thereof is 0.6kB (kilobytes), the vocalization time of "much of a" is 0.4 seconds and the capacity thereof is 2.4kB (kilobytes), and the

vocalization time of "problem" is 0.3 seconds and the capacity thereof is 1.8kB (kilobytes), and the vocalization time of the entire segment 621 is 2.0 seconds and the capacity thereof is 12kB (kilobytes).

[0077] Furthermore, Fig. 10 is a table for explaining the recording contents in a directory region in the case shown in Fig. 8 and Fig. 9. In this table, the directory region is formed by  $9 \times 3 = 27$  bytes (B) per segment. The speech sound data sequences A, B, and C correspond to the speech sound data sequences A, B, and C of Fig. 8, respectively. Furthermore, C with 1 byte indicates the attribute, and C=0 means the speech sound data sequence A, and C=64 means the speech sound data sequence B. Furthermore, C=128 or 129 means the speech sound data sequence C, and in particular, when C=129, that is, in the case of "10000001" in bit expression (8 bits), it indicates the same segment for explanation as the previous segment (indicates that it belongs to the same segment group for explanation of the speech sound data sequence C, and this correspond to the case of the segment 801 or 802 of Fig. 4(e)).

[0078] M, S, and B (each has one byte) of position data are parameters indicating positions on a CD-ROM that is a standard in the industrial world. Namely, M indicates minutes, S indicates seconds, and B indicates blocks. 1 block is 2,048

bytes, and 75 blocks correspond to one second. Therefore, the maximum values are M=59, S=59, and B=74. The next SB with 2 bytes indicates starting bytes, and the next LLL with 3 bytes indicates the length of the entirety of each segment. The reason for using minutes and seconds for parameters for indicating positions is that a CD-ROM has been originally developed for music, and the recorded position is expressed as time from the beginning. Therefore, in a case where a CD-ROM is employed as the abovementioned speech sound data recording medium, the minutes and seconds have no relationship with the reproduction period, and are data that only shows a recorded position on the recording medium.

[0079] As a result, for example, the segment 621 "It's not much of a problem." of the speech sound data sequence A is recorded as English speech sound data spoken by a native speaker with a length of 6,000 bytes from the 826th byte of 0 minute 11 seconds and 3 blocks, and the corresponding segment of the speech sound data sequence B is recorded as English spoken by the native speaker selectively extended so as to have a length of 17,400 bytes from the 2,026th byte of 0 minute 11 seconds and 3 blocks, and the segment group of the speech sound data sequence C is recorded as explanation in Japanese with a length of 72,000 bytes from the 1,282nd byte of 0 minute 11 seconds



and 6 blocks. The segment numbers of 621 and 622, etc., do not exist on the memory, and correspond to the addresses of the segments. Furthermore, the recorded position identification data indicating the relationship of the segments is contained in this directory region.

[0080] Furthermore, in detail, from the recorded contents in the directory region shown in Fig. 10, in a region from the 826th byte of 0 minute 11 seconds and 3 blocks of this speech sound data recording medium to  $826+6,000-1 = 6,825$ th byte, a speech sound data sequence of the segment number 621 and attribute  $C=0$ , that is, data corresponding to "It's not much of a problem." spoken by a native speaker is recorded. In a region from the 2,026th byte to  $2,026+17,400-1 = 19,425$ th byte of 0 minute 11 seconds and 3 blocks of this speech sound data recording medium, a speech sound data sequence of the segment number 621 and attribute  $C=64$ , that is, speech sound data selectively extended is recorded. Furthermore, in a region from the 1,282nd byte to  $1,282+72,000-1 = 73,281$ st byte of 0 minute 11 seconds and 3 blocks of the speech sound data recording medium, a speech sound data sequence of the segment 621 and attribute  $C=128$ , that is, data corresponding to explanation in Japanese is recorded.

[0081] By thus providing the directory region shown in Fig.

10, the respective speech sound data sequences shown in Fig. 8 can be recorded with the reproduction period and capacity shown in Fig. 9.

[0082] Next, data on the segments 621 and 622 is recorded on the header part of, for example, the variable-length segment shown in Fig. 11(a). This header part is a 32-byte region composed of, as shown in Fig. 11(b): a 1-byte region (1B) for showing existence of text data and image data from the header, a 5-byte region (5B) that is prepared for the speech sound data sequence A and is composed of 1-byte data showing the data sequence type (data for distinction between the speech sound data sequences A and B, etc.), 3-byte data showing the data length, and preliminary 1-byte data; a 5-byte region (5B) that is prepared for the speech sound data sequence B and composed of 1-byte data indicating the data sequence type, 3-byte data indicating the data length, and preliminary 1-byte data; a 4-byte region (4B) that is prepared for the speech sound data sequence C and is composed of 1-byte data showing the data sequence type and 3-byte data showing the data length; a 4-byte region (4B) that is prepared for the text data sequence D and is composed of 1-byte data showing the data sequence type and 3-byte data showing the data length; a 6-byte region (6B) that is prepared for the same text data sequence D and is composed

of 3-byte data showing the address and 3-byte data showing the data length; a 4-byte region (4B) that is prepared for another data sequence (Type E) such as the third speech sound data sequence; and a preliminary 3-byte region (3B).

[0083] Next, the constructions of the speech sound data reproducing method and unit relating to the invention are described with reference to Fig. 12 through Figs. 14.

[0084] First, Fig. 12 is a perspective view showing the entire construction of a reproducing unit for realizing the speech sound data reproducing method relating to the invention. As seen in this figure, the speech sound data recording medium is, for example, a CD-ROM that can be reproduced by a portable CD player (reproducing unit main body 200), and this reproducing unit main body 200 is remotely controlled by a handset 80 connected by a cord. This handset 80 is provided with a display part 210 such as a liquid crystal display (LCD), etc., for displaying at least a segment number of a segment that is being reproduced, and a button group 240 for various controls. An operator can listen to the speech sound data reproduced by the reproducing unit main body 200 through an earphone 130.

[0085] Fig. 13 is a block diagram showing the construction of the reproducing unit shown in Fig. 12. As shown in this figure,

the CD-ROM as the speech sound data recording medium 15 is set into a reproducing mechanism 205. The reproducing mechanism 205 is connected to the CPU 50 via a disk interface (I/F) 30 and a bus 40. To the bus 40, a ROM 60 of, for example, 32 kilobytes (kB) for storing programs and a RAM 70 of, for example, 256 kilobytes for temporarily storing directories and speech sound data sequences are connected. Furthermore, to the bus 40, a handset interface (I/F) 90 for data exchange with the handset 80 for manual operation and a D/A converter 12 connected to an external terminal 110 and the handset 80 via a speech sound outputting amplifier (AMP) 100 are connected.

[0086] Fig. 14(a) and Fig. 14(b) are drawings for explaining the memory assignment conditions of the ROM 60 and the RAM 70, respectively. As shown in Fig. 14(a), the 32kB ROM 60 stores programs. On the other hand, as shown in Fig. 14(b), to the RAM 70, a buffer (corresponding to 50 blocks) of  $(50+50) = 100$  kilobytes, a directory of  $(75+75) = 150$  kilobytes, and a system area of 6 kilobytes are assigned. Therefore, in the RAM 70, speech sound data sequences corresponding to 50 blocks are always stored, and a directory corresponding to 150 kilobytes/ $27 \approx 5,555$  segments is stored (only the speech sound data sequence A corresponds to approximately 30 minutes).

[0087] In the abovementioned detailed examples, a CD-ROM is

used as the speech sound data recording medium, and a typical CD-ROM has a capacity of 552 megabytes (MB). In a CD-ROM, for expressing an address, minutes, seconds, and blocks are used as units. One block has 2,048 bytes, 75 blocks correspond to 1 second, and 60 seconds correspond to one minute, so that the maximum value of the address in the CD-ROM is 59 minutes 59 seconds and 74 blocks. On the contrary, the capacity of this CD-ROM is  $2,048 \times 75 \times 60 \times 60 = 552.96$  megabytes. In this capacity, a capacity corresponding to 2 seconds from the beginning is for the CD-ROM format and cannot be used by a user, so that the maximum capacity is accurately 552.6528MB. Furthermore, when a directory region is assigned to a capacity corresponding to 20 seconds from the beginning, a 3-megabyte directory capacity can be secured in the CD-ROM.

[0088] Furthermore, in the abovementioned examples of the invention, a singular sale of speech sound data recording software (program that enables a personal computer, etc., to execute the abovementioned recording method, or a recording medium containing this program recorded), an exclusive recording unit, and a user's manual, or a sale of a combination of these, a singular sale of a speech sound data recording medium, and furthermore, a sale of this speech sound data recording medium, reproducing software (including programs

that can be executed by a personal computer, etc., or a recording medium containing these programs recorded), an exclusive reproducing unit, and a user's manual, or a sale of a combination of these are possible.

[0089]

[Effects of the Invention] As mentioned above, in the invention, by changing (emphasizing or attenuating) the amplitude or the number of waves (increasing the number for extension of the reproduction period or reducing the number for contraction of the reproduction period) at an arbitrary part of a plurality of frequency components divided from a first speech sound data sequence sampled in a first period, corrected sine wave data is generated, and sine wave data of the respective frequency components are summed to newly synthesize a second speech sound data sequence, and the second speech sound data sequence is recorded on a predetermined recording medium. The desired speech sound data sequence thus recorded has an effect in that it can be reproduced as speech sound which has a reproduction period extended or contracted at an arbitrary part or is emphasized or attenuated at an arbitrary part without changes in frequency.

[0090] Furthermore, this invention can be combined with the technique disclosed in Japanese Patent No. 2581700, and by

separately preparing speech sound data made by extending and/or emphasizing vocal sound spoken by a native speaker at an arbitrary part according to variable-length sections divided at vocalization pauses, it becomes possible for novice learners to repeatedly reproduce and listen to speech sound that could not heard by them, and also to emphasize/extend and listen to a part of speech sound that is hardly heard. Furthermore, for expert learners, by separately preparing speech sound data contracted and/or attenuated at an arbitrary part, more positive learning becomes possible in conjunction with reproduction of vocal sound spoken by a native speaker.

[BRIEF DESCRIPTION OF THE DRAWINGS]

[Fig. 1] A block diagram for explaining the recording operation of speech sound data relating to the invention.

[Fig. 2] A table showing an example of frequency components (channels) divided from sampled input speech sound data.

[Fig. 3] A drawing for explaining a basic form of a speech sound spectrum.

[Figs. 4] Drawings for conceptually explaining various data including speech sound data sequences to be recorded on the speech sound data recording medium relating to the invention.

[Fig. 5] A drawing showing the entire construction of a peripheral device for realizing the speech sound data recording

method relating to the invention.

[Fig. 6] A flowchart (1) for explaining the speech sound data recording method relating to the invention.

[Fig. 7] A flowchart (2) for explaining the speech sound data recording method relating to the invention.

[Fig. 8] A drawing for explaining the respective speech sound data sequences and their recording contents on the speech sound data recording medium applied for English conversation self-study.

[Fig. 9] A table for explaining the relationship between time and capacity per segment of each speech sound data sequence shown in Fig. 8.

[Fig. 10] A table for explaining recording contents (including recorded position identification data) of a directory region in the speech sound data recording medium shown in Fig. 8 and Fig. 9.

[Figs. 11] Drawings showing the construction of variable-length segments to be recorded on the speech sound data recording medium relating to the invention.

[Fig. 12] A perspective view showing the entire construction of a reproducing unit for realizing the reproducing method for the speech sound data recording medium relating to the invention.



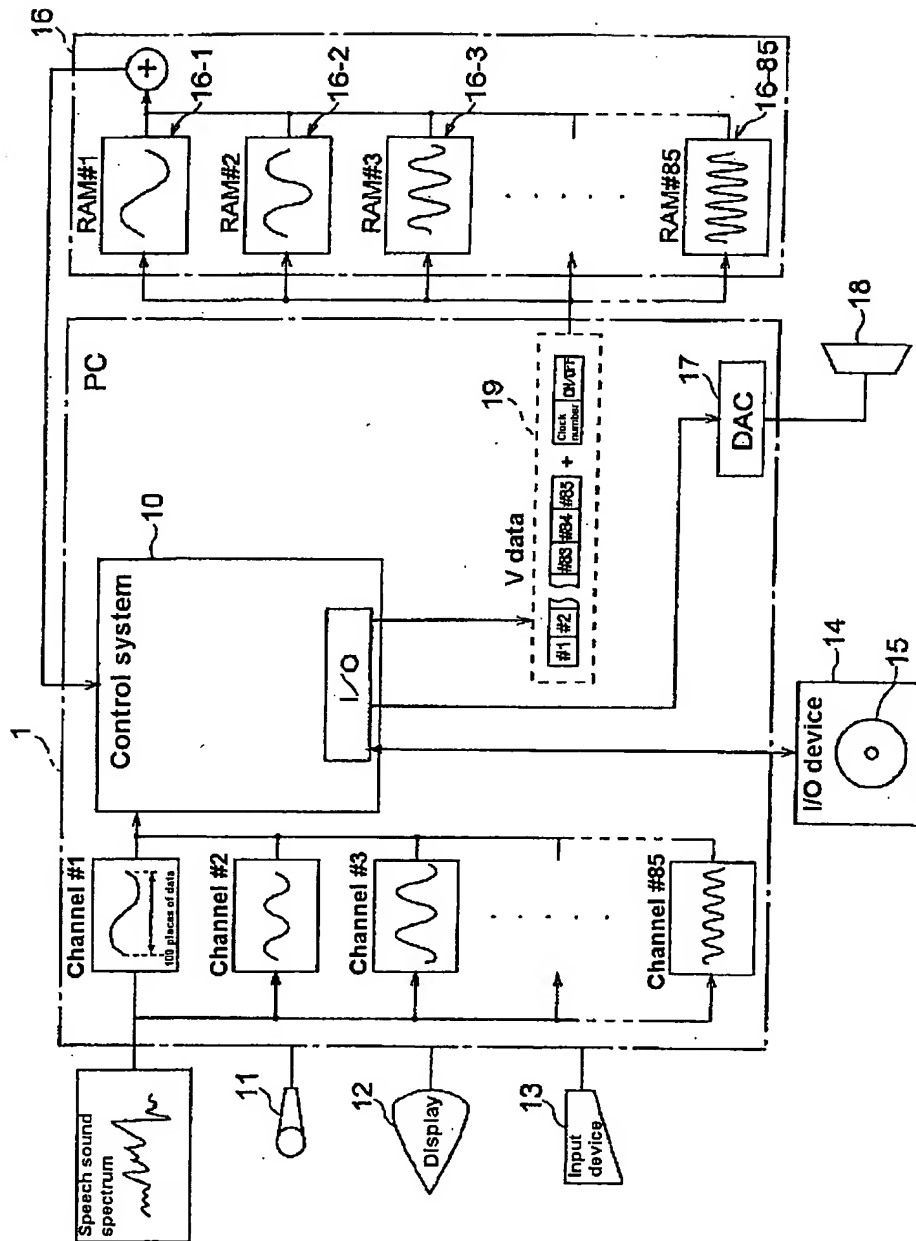
[Fig. 13] A block diagram showing the construction of the reproducing unit shown in Fig. 12.

[Figs. 14] Drawings for explaining the memory assignment conditions of the ROM and RAM shown in Fig. 13.

[Description of Symbols]

1...PC, 10...control system, 14...I/O device, 15...speech sound data recording medium, 19...V data, 16-1 through 16-85...sine wave data generating circuit, 17, 175...DAC, 18, 177...speaker, 173...adder

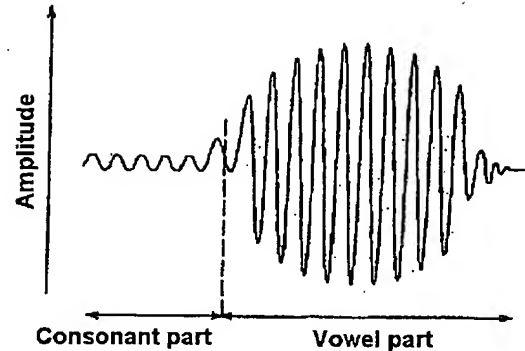
# Fig.1



# Fig.2

Channel	Low pass f	High pass f	Center f	Scale	Clock/wave (64, 128)	Clock number	Wave number	Error (ppm)
#1	75.57	80.08	77.78	D#	566.98	567	1	28.8
#2	80.06	84.82	82.41	E	535.13	535	1	242
#3	84.82	89.87	87.31	F	505.10	505	1	192
#4	89.87	95.21	92.50	F#	476.75	477	1	510
#5	95.21	100.9	98.00	G	450.00	450	1	0
#6	100.9	106.8	103.8	G#	424.88	425	1	340
#7	106.8	113.2	110.0					

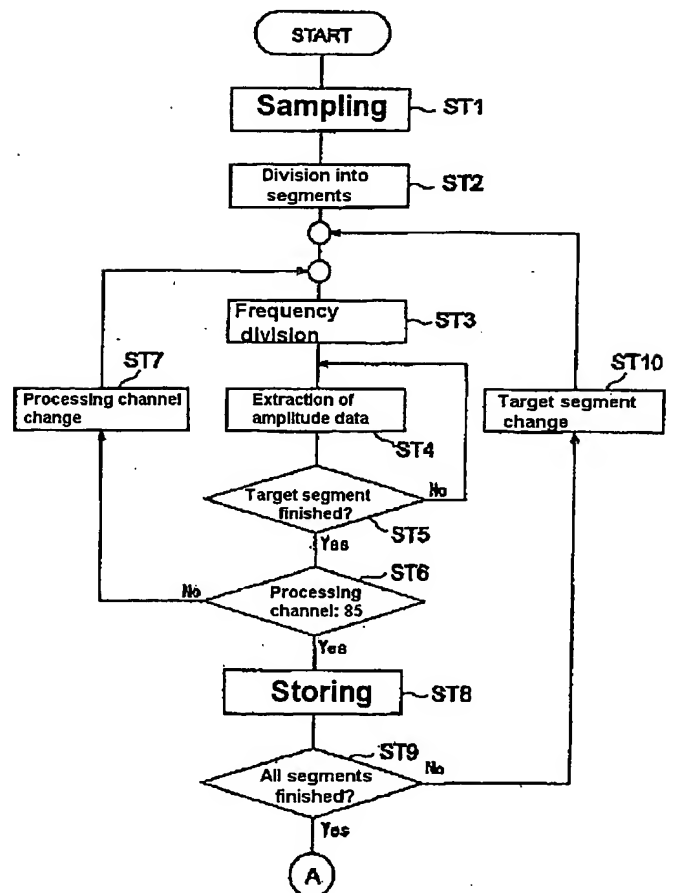
# Fig.3



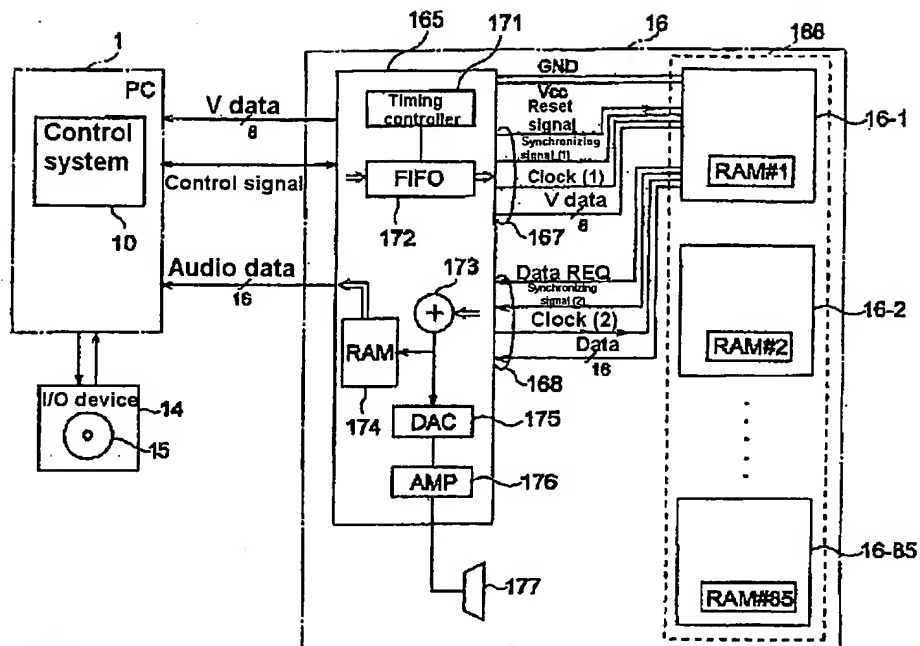
# Fig.4

- Segment 621
- (a) It's not much of a problem.
- Segment 622
- (b) I'd second that.
- Segment 701
- (c) Yes!
- Segment 799
- (d) I'm gonna go fishing with my brother.
- Segment 801
- (e) I'm gonna go fishing with my brother,
- Segment 802
- because he'd bought a new rod yesterday.

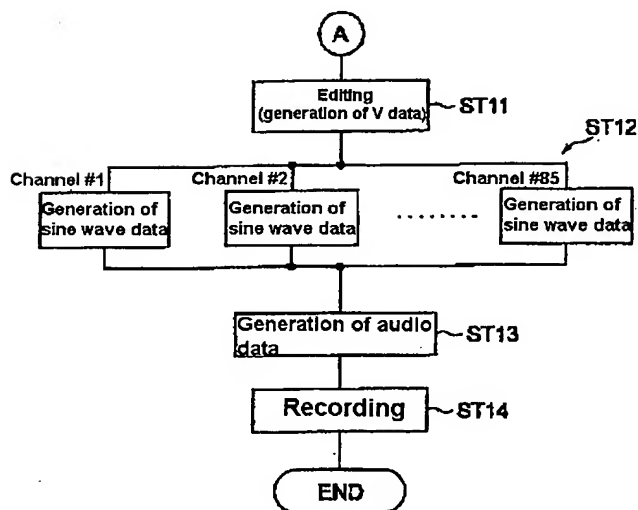
# Fig.6



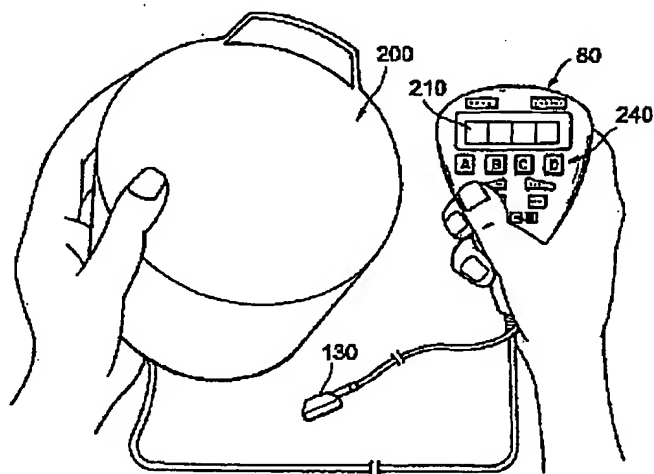
# Fig.5



# Fig.7



# Fig.12



# Fig.9

Segment		621	622	
Time (sec)	A	2.0	1.3	
	B	3.9	2.6	
	C	12	25	
Capacity (KB)	A	12	7.8	
	B	23.4	15.6	
	C	72	150	

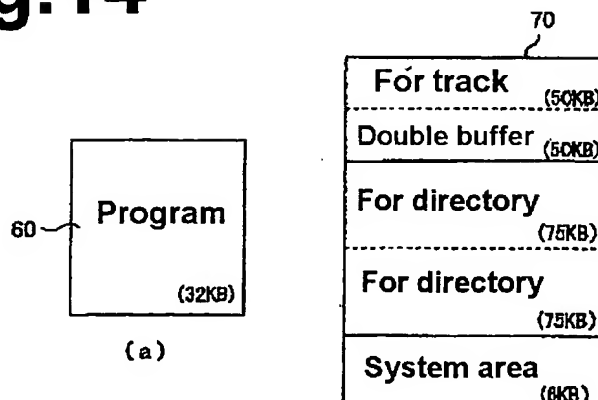
# Fig.8

Recording contents	
A	<div> <div>"It's not much of a problem."</div> <div>I'd second that."</div> </div> <div> <div>621</div> <div>622</div> </div>
B	<div> <div>It's...not...much...of...a...problem...</div> <div>I'd...second...that...</div> </div> <div> <div>621</div> <div>622</div> </div>
C	<div>(621)</div> <div>"'It' indicates the previous 'something' and 'not much of a problem' is an idiom meaning 'no problems' or 'no worries'."</div>
	<div>(622)</div> <div>Next, 'I'd' is an abbreviation of 'I would' and 'I could' can also be abbreviated to 'I'd' in the same manner. 'Second' means to 'support' or to 'agree with' and 'I'd second that' means 'I agree with that'."</div>

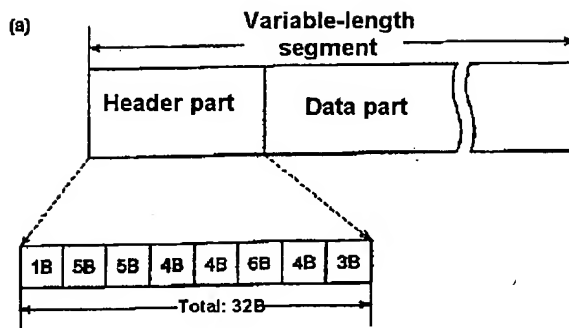
# Fig.10

Data sequence	Segment	Category	Position data			Starting byte	Segment quantity (byte)
	SS	C	M	S	B	SB	LLL
A							
A	621	0	0	11	3	826	6,000
B		64	0	11	3	2,026	17,400
C		128	0	11	6	1,282	72,000
A	622	0	0	11	49	2,018	4,800
B		64	0	11	50	570	12,600
C		128	0	11	52	1,274	150,000

# Fig.14



# Fig.11



(b)

Header (Total:32B)	Data sequence	Contents
1B		Existence of text data and image data (1B)
5B	A	Data sequence type (1B), data length (3B), preliminary (1B)
5B	B	Data sequence type (1B), data length (3B), preliminary (1B)
4B	C	Data sequence type (1B), data length (3B)
4B	(D)	Data sequence type (1B), data length (3B)
6B	D	Address (3B), data length (3B)
4B	E	
3B	Preliminary	

# Fig.13

